



**HAL**  
open science

## Could genomic selection methods be efficient to detect QTLs? Application in French dairy cattle

Carine Colombani, Pascal Croiseau, Chris Hoze, Sebastien Fritz, François Guillaume, Didier Boichard, Andres Legarra, Vincent Ducrocq, Christèle Robert-Granié

► **To cite this version:**

Carine Colombani, Pascal Croiseau, Chris Hoze, Sebastien Fritz, François Guillaume, et al.. Could genomic selection methods be efficient to detect QTLs? Application in French dairy cattle. QTLMAS, May 2011, Rennes, France. pp.30. hal-01190269

**HAL Id: hal-01190269**

**<https://hal.science/hal-01190269>**

Submitted on 3 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# The 15<sup>th</sup> Quantitative Trait Locus and Marker Assisted Selection Workshop

Rennes – Agrocampus Ouest , Brittany



**The book of abstracts**

**15th QTL–MAS Workshop**

**The 15<sup>th</sup> QTL-MAS workshop is organized by INRA, the French “National Institute for Agricultural Research” in collaboration with Agrocampus Ouest.**

The core of the Organizing Committee consists of:

Olivier Demeure

Jean Michel Elsen

Olivier Filangi

Christine Gourbe

Pascale Le Roy

The Workshop is organized with support from:

Anne-Yuna Blum

Colette Désert

Cécile Duby

Yoannah Francois

Magalie Houée

Thomas Obadia

Marion Ouedraogo

Hélène Romé

Simon Teyssedre

Xiaoqiang Wang

This book of abstracts has been edited by the Organizing Committee.

## HONORARY PATRONAGE

15<sup>th</sup> QTL-MAS 2011 Workshop has been organized under the honorary patronage of :



**French National Institute for  
Agricultural Research**

<http://www.international.inra.fr/>



**AGROCAMPUS OUEST is a higher  
education and research public Institute  
of the French Ministry of Food,  
Agriculture and Fisheries**

<http://www.agrocampus-ouest.fr>

## PARTNERS

The following sponsors have provided financial support to  
the 15th QTL-MAS workshop



### Region Bretagne

5 rue Martenot  
35000 Rennes

Tél : 02 99 27 18 30/31

Fax : 02 99 27 18 06

Courriel : [conseil.culturel@region-bretagne.fr](mailto:conseil.culturel@region-bretagne.fr)

Website : <http://www.bretagne.fr/>



### Rennes Métropole

Hôtel de Rennes Métropole  
4, avenue Henri-Fréville  
CS 20723  
35207 rennes cedex

Téléphone : 02.99.86.60.60

Fax : 02.99.86.61.61

Website : <http://www.rennes-metropole.fr/>



**Cobb Vantress**

Cobb Europe Ltd  
Oyster House, Severalls Lane, Colchester  
Essex CO4 9PD, UK  
Tel: +44 1206 835835

**Website :** <http://www.cobb-vantress.com/>



**Grimaud Frères Sélection**

La Corbière  
49450 ROUSSAY  
FRANCE

Tél. +33 (0)2 41 70 36 90  
Fax : +33 (0)2 41 70 31 67

**Website :** <http://www.grimaudfreres.com/>



**Hubbard**

Hubbard S.A.S.  
Mauguerand, Le Foeil - P.O. Box 169  
22800 Quintin  
FRANCE

TEL. +33-2.96.79.63.70  
FAX +33-2.96.74.04.71  
Email: [contact.emea@hubbardbreeders.com](mailto:contact.emea@hubbardbreeders.com)  
**Website :** <http://www.hubbardbreeders.com/>



**Hendrix Genetics**

Research & Technology Centre  
Villa "de Körver"  
Sporstraat 69  
P.O.Box 114  
5830 AC Boxmeer  
The Netherlands-EU

T +31 485 801981  
F + 31 485 801982

[Info@hendrix-genetics.com](mailto:Info@hendrix-genetics.com)

**Website :** <http://www.hendrix-genetics.com/>

# Program

## Thursday 19/05/11

8h30-9h00

*Registration*

9h00-9h30

*Opening and Introduction*

### **Session 1 . Methods for QTL detection and Genetic Architecture Analysis (1/4)** **Chair : Tomasz Strabel**

9h30-10h10	<b>A Hidden Markov Model Combining Linkage and Linkage Disequilibrium Information for Haplotype Reconstruction and Quantitative Trait Locus Fine Mapping (invited)</b>	<b>Tom Druet</b>
10h10-10h25	Inferring haplotypes and parental genotypes in larger full sib-ships and other pedigrees with missing genotype data	Carl Nettelblad
10h25-10h40	Genome-wide analysis in a Chinese x Western chicken F2 intercross	Zheya Sheng
10h40-10h55	Transcriptome profiling reveals interaction between two QTL for fatness in chicken	Yuna Blum

*COFFEE BREAK*

### **Session 1 . Methods for QTL detection and Genetic Architecture Analysis (2/4)** **Chair : Suzanne Rowe**

11h25-12h05	<b>Genetic dissection of susceptibility for Crohn's disease (invited)</b>	<b>Jean-Pierre Hugot</b>
12h05-12h20	QTL detection for survival related traits in atlantic salmon	Francois Besnier
12h20-12h35	Genetic and genomic study of the rainbow trout <i>Oncorhynchus Mykiss</i> response to a salinity change	Yvan Lebras
12h35-12h50	A comparison of regression models for testing QTL/eQTL co location	Xiaoqiang Wang
12h50-13h05	A simulation-based approach for controlling the false positive rate in genome-wide association studies	Marcin Kierczak

*LUNCH*

### **Session 1 . Methods for QTL detection and Genetic Architecture Analysis (3/4)** **Chair : François Besnier**

14h15-14h30	Mapping QTL from QTL-MAS workshop data using four methods	Naveen Kumar Kadri
-------------	---	--------------------



15<sup>th</sup> QTL-MAS Workshop, May 19-20, 2011 , Rennes, France

14h30-14h45	Bayesian Variable Selection to identify QTL affecting a simulated quantitative trait	Anouk Schurink
14h45-15h00	Estimation of breeding values and detection of QTL in QTL-MAS 2011 dataset using GBLUP and Bayesian approaches	Marcin Pszczola
15h00-15h15	Genomic Breeding Value Prediction and QTL Mapping of QTLMAS2011 Data Using Bayesian Methods	Dorian Garrick

*COFFEE BREAK*

**Session 1 . Methods for QTL detection and Genetic Architecture Analysis (4/4)**

**Chair : Jean Pierre Bidanel**

15h45-16h00	Variance Controlling Genes: Significant Contributors to The Missing Genetic Variation for Complex Traits	Xia Shen
16h00-16h15	Use of planes in genotype space to detect high order epistasis	Mats Pettersson
16h15-16h30	A problem of locating multiple interacting QTL by logic regression	Magdalena Malina
16h30-16h45	qtl.outbred: Interfacing outbred line cross data with the R/qtl mapping software	Ronald M. Nelson
16h45-17h00	AnnotQTL: a new tool to gather functional and comparative information on a genomic region	Olivier Demeure
17h00-17h15	GPU accelerated QTLMap	Guillaume Chapuis

17h30

*Departure for Rennes City Center*

18h00

*Cocktail at Rennes City Hall offered by Rennes Métropole*

19h00-20h45

*Visit of the Brittany Parliament (4 groups)*

21h00

*Dinner at the "Taverne de la Marine", Place de Bretagne*

## Friday 20/05/11

### Session 2 . Methods for Genomic Evaluation (1/2)

Chair : Rodolfo Cantet

9h00-9h40	<b>Strategies to improve the accuracy of genomic predictions (invited)</b>	<b>Dorian Garrick</b>
9h40-9h55	Could genomic selection methods be efficient to detect QTL?	Carine Colombani
9h55-10h10	Linear models for breeding values prediction in haplotype-assisted selection	Anna Mucha
10h10-10h25	Effect of the prior distribution of SNP effects on the estimation of total breeding value	Valentina Riggio
10h25-10h40	Approximate Bayesian prediction of genomic breeding values and variances using INLA	Patrik Waldmann

*COFFEE BREAK*

### Session 2 . Methods for Genomic Evaluation (2/2)

Chair : Frédéric Hospital

11h10-11h50	<b>Genomic selection: implementation in dairy cattle (invited)</b>	<b>Vincent Ducrocq</b>
11h50-12h05	A simple two step Bayesian approach for genomic prediction of breeding values	Mohammad Shariati
12h05-12h20	Genomic Analyses for the common dataset of the 15th QTL-MAS Workshop	Chonglong Wang
12h20-12h35	Alternative strategies for selecting subsets of predicting SNPs by LASSO-LARS procedure	Graziano Usai
12h35-12h50	Genomic selection using regularized linear regression models: ridge regression, lasso and elastic net and their extensions	Joseph O Ogutu
12h50-13h15	<i>Poster session</i>	

*LUNCH*

### Session 3 . Comparative analysis of submitted QTLMAS2011 Results

Chair : Olivier Filangi

14h15-14h35	All the truth about the QTLMAS2011 Data	Jean Michel Elsen
14h35-14h55	Comparison of submitted methods for QTL identification in the QTLMAS2011 Data set	Olivier Demeure
14h55-15h15	Comparison of submitted methods for GEBV estimation from QTLMAS2011 Data set	Pascale Le Roy
15h15-15h45	Closing Remarks	

# Abstracts

# **A hidden Markov model combining linkage and linkage disequilibrium information for fine-mapping studies**

Tom Druet\* and Michel Georges

Unit of Animal Genomics, GIGA-Research and Department of Animal Production, Faculty of Veterinary Medicine, University of Liège, B-4000 Liège, Belgium

\*Presenting author: Tom Druet, email: tom.druet@ulg.ac.be

Accurate reconstruction of haplotypes from diploid marker data (phasing) is of importance for many kinds of genetic analyses including mapping of trait loci, prediction of genomic breeding values, and identification of signatures of selection. In human genetics, phasing most often exploits population information (linkage disequilibrium), while in animal genetics the primary source of information is familial (Mendelian segregation and linkage). Recently, we have developed a method that takes advantage of both sources of information by extending hidden Markov Models limited to linkage disequilibrium information. Our new approach has been shown to improve accuracy of allele phasing as well as imputation of missing genotypes. Reconstructed haplotypes are assigned to hidden states which correspond to clusters of genealogically related chromosomes (i.e., ancestral haplotypes). In this talk, we show how these ancestral haplotypes can be used directly to fine-map QTL and genetic disorders. Furthermore, the method is computationally effective at handling large data sets, such as those obtained from high density SNP panels.

# Inferring haplotypes and parental genotypes in larger full sibships and other pedigrees with missing genotype data

Carl Nettelblad<sup>1\*</sup>

<sup>1</sup>Division of Scientific Computing, Department of Information Technology, Uppsala University, Box 337, SE-75105, Uppsala, Sweden

\*Presenting author: Carl Nettelblad, email: carl.nettelblad@it.uu.se

**Background.** In many contexts, pedigrees for individuals are known even though not all individuals have been fully genotyped. In one extreme case, the genotypes for a set of full siblings are known, with no knowledge of parental genotypes. We propose a method for inferring haplotypes and genotypes for all individuals, even those with missing data, in such pedigrees, allowing a multitude of classic and recent methods for linkage analysis to be used.

**Results.** By artificially removing the founder generation genotype data from the dataset simulated for the 15th QTL-MAS workshop, the quality of reconstructed genotypes in that generation can be verified. For the full structure of repeated matings with 15 offspring per mating, 10 dams per sire, 99.86% of all markers were phased correctly. When reduced to only 5 full-sib offspring in a single sire-dam mating, the corresponding percentage is 92.62%, which compares favorably with 89.28% from the Merlin package [1]. That package is also unable to handle more than approximately 10 sibs, as the number of states tracked rises exponentially with family size.

**Conclusions.** Our method is able to reconstruct genotypes for parents when such data is missing, as well as haplotypes for all individuals, in an extension to our previous work where parental genotypes were needed. Tracking of haplotype and allele origin is important in all applications where the marker set might not directly reflect genotype variation influencing traits. The cnF2freq codebase, which is in a state of active development, is available under a BSD-style license.

1. Abecasis G, Cherny S, Cookson W, Cardon L: Merlin - rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics* 2001, 30:97–101.

## **Genome-wide analysis in a Chinese x Western chicken F<sub>2</sub> intercross**

Zheya Sheng\*, Mats Pettersson, Xiaoxiang Hu, Örjan Carlborg, Ning Li

\*Presenting author: Zheya Sheng, email: Zheya.Sheng@slu.se

Broiler meat consumption is increasing rapidly in China today. To meet the demand, new efficient breeding programs are needed to balance requirements of meat quality and production efficiency. One of the challenges faced by breeders is to select the most appropriate breeds to start selection from. By utilizing genomics technology it is possible to characterize the genetic architecture of important traits in multiple breeds and in this way allow a more informed choice in the breeding process.

China has a large number of native chicken breeds, each with their own unique qualities. Meat quality has been a key selection trait for most breeds and consequently they have more desirable meat quality characteristics than the western commercial breeds. On the other hand, the western commercial broilers have a higher production efficiency. To characterize the genetic differences between western and Chinese breeds we produced an F<sub>2</sub> intercross between a Chinese indigenous breed, Huxu, and an imported commercial breed, A03.

We genotyped nearly 600 F<sub>2</sub> individuals using a 60K SNP-chip and recorded more than 100 traits including body weight, meat-quality related and appearance traits. A genome-wide association mapping was done on phenotypes that were pre-corrected for fixed and random environmental effects using REML. The family structure was accounted for by including a genetic kinship matrix as a random effect. Significance testing was based on bonferroni corrected p-values and genomic control. Multiple significant signals were detected for the analyzed traits. QTL analysis, including genome-scans for epistasis, is also going to be performed in addition to the GWAS analysis to make use of the F<sub>2</sub> pedigree structure in the mapping. This work will give new insights to the genetic differences between Chinese and Western breeds for a wide range of production and quality traits and is an important integral component of the development of future chicken breeding programs in China.

## Transcriptome profiling reveals interaction between two QTL for fatness in chicken

Yuna Blum<sup>1,2,3</sup>, Guillaume Le Mignon<sup>1,2,4</sup>, Frédérique Pitel<sup>5</sup>, David Causeur<sup>3</sup>, Olivier Demeure<sup>1,2</sup>, Pascale Le Roy<sup>1,2</sup>, Sandrine Lagarrigue<sup>1,2</sup>

<sup>1</sup>INRA, UMR598, Génétique Animale, IFR140 GFAS, 35000 Rennes, France

<sup>2</sup>Agrocampus Ouest, UMR598, Génétique Animale, IFR140 GFAS, 35000 Rennes, France

<sup>3</sup>Agrocampus Ouest, Applied Mathematics Department, 35000 Rennes, France

<sup>4</sup>ITAVI, F-75008, Paris, France

<sup>5</sup>INRA, UR444, Laboratoire de Génétique Cellulaire, 31326 Auzeville, France

\*Presenting author: Yuna Blum, email: yuna.blum@rennes.inra.fr

In the “genetical genomics” field, different approaches are proposed to improve characterization of QTL regions for various traits combining QTL detection and transcriptome profiling. The most common approach is to identify genes whose eQTL colocalize with QTL of interest, providing new functional hypothesis about the QTL causative mutation. A second approach used only by Schadt et al (2003) and our laboratory (Le Mignon et al (2009)) consists in performing linkage analysis for the trait using only some of the animal subgroups of a F2 population generated on the basis of their transcriptome profiles. Such an approach can refine some QTL and detect new ones.

This approach was applied to hepatic transcriptome profiles for 45 offspring of a chicken known to be heterozygous for a QTL for abdominal fatness (AF) on GGA5 at 168cM. 688 gene expressions significantly correlated to the AF trait were obtained using a recent method taking into account the hidden dependence structure independent of the trait of interest as the population structure or microarray expression artefacts (Friguet *et al*, 2009, Blum *et al*, 2010). A hierarchical clustering analysis using these 688 genes distinguished five subtypes for fatness among the 45 birds. After removing a subtype of 7 lean birds, linkage analysis revealed another QTL on the GGA5 at ~102cM. Interestingly, these 7 animals presented the same paternal haplotypes for the QTL at 168cM, suggesting that the two QTL are in interaction. We showed by different approaches (ANOVA, linkage analysis) a significant interaction between the two QTL. These results illustrate the power of the approach: transcriptome data allows separating a population into genetically homogenous phenotype subgroups, revealing the complexity of the genetic regulation of a complex trait.

## QTL detection for survival related traits in atlantic salmon

Francois Besnier<sup>1</sup>\*, Kevin Glover<sup>1</sup>, Øystein Skaala<sup>1</sup>

<sup>1</sup>Institute of Marine Research, PO Box 1870, N-5817 Bergen, Norway.

\*Presenting author: Francois Besnier, email: francois.besnier@imr.no

**Background.** Fish farming has become an economic successful industry in several countries. Due to the high fecundity of fish species, large numbers of offspring can be produced from relatively small numbers of brood stock. This situation led to a rapid differentiation of the farmed populations under the effect of selective breeding for production related traits. Due to this fast differentiation between farm and wild populations, the high numbers of farmed individuals present in sea cages may represent a threat for the wild populations through escapement and competition for resources such as food and spawning areas, as well as through gene flow from escapees into wild gene pools. Along with the continued expansion of Atlantic salmon farming, concern about the genetic and ecological impacts on wild salmon populations is growing.

**Method.** To determine the relative survival and performance of hybrids from farm-escapes and wild salmon compared with pure wild and farmed fishes in natural environment, 49 families of either wild, farm or hybrid lineage were generated. Eggs from these families were planted in years 2004 and 2005, in the Guddalselva river, on the West coast of Norway. During the following years after the introduction of salmon eggs in the river, a trap was installed downstream in order to catch all salmon during their migration from fresh water to the sea (smolt life stage). All fishes collected in the trap were phenotyped for length and body weight, and fin-clipped for future genotyping and family assessment.

**Results.** A total of 2500 salmon from cohort 2004 and 2005 were captured in the trap and later genotyped for 272 SNP markers placed at regular intervals along the 29 linkage groups of the salmon genome. All linkage groups were subsequently scanned for QTL affecting the length (l), weight (w) or condition factor ( $cf=w/(l^3)$ ) via variance component linkage mapping. Preliminary results show that several regions contain QTL affecting salmon's weight and condition factor. Moreover these region seem to overlap with several QTL found by Houston et al. (2009), and Baranski et al. (2010).

**Perspective.** These results points the Guddalselva experimental data as an outstanding material to study the relative performance of wild vs. farm and hybrid salmon in natural condition, and assess the genetic polymorphism underlying these differences. Further analyses will use more powerful approach for QTL detection such as LDL mapping (Hayes et al. 2006, Meuwissen and Goddard 2001), and will focus on scanning for QTL affecting survival.

*BJ Hayes, A Gjuvslund and S Omholt. (2006). Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males. Heredity 97, 19–26*  
*Meuwissen THE, Goddard ME (2001). Prediction of identity by descent probabilities from marker-haplotypes. Gen Sel Evol 33: 605–634.*

*R. D. Houston, S. C. Bishop, A. Hamilton, D. R. Guy, A. E. Tinch, J. B. Taggart, A. Derayat, B. J. McAndrew and C. S. Haley (2009) Detection of QTL affecting harvest traits in a commercial Atlantic salmon population. Animal Genetics, 40, 753–755*  
*Matthew Baranski, Thomas Moen and Dag Inge Våge (2010). Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (Salmo salar). Genetics Selection Evolution 2010, 42:17*



## Genetic and genomic study of the rainbow trout *Oncorhynchus Mykiss* response to a salinity change

Yvan Le Bras<sup>1\*</sup>, Nicolas Dechamp<sup>2</sup>, Jerome Montfort<sup>1</sup>, Aurélie Le Cam<sup>1</sup>, Grégory Guernec<sup>1</sup>, Olivier Filangi<sup>3,4</sup>, Francine Krieg<sup>2</sup>, René Guyomard<sup>2</sup>, Mekki Boussaha<sup>2</sup>, Henk Bovenhuis<sup>5</sup>, Tom G. Pottinger<sup>6</sup>, Edwige Quillet<sup>2</sup>, Patrick Prunet<sup>1</sup>, Pascale Le Roy<sup>3,4</sup>

<sup>1</sup>INRA, UR1037 SCRIBE (Rennes)

<sup>2</sup>INRA, UMR1313 GABI (Jouy-en-Josas)

<sup>3</sup>INRA, UMR0598 Génétique Animale (Rennes)

<sup>4</sup>Agrocampus Ouest, UMR0598 Génétique Animale (Rennes)

<sup>5</sup>Animal Breeding and Genetics Group, Wageningen university (Wageningen, The Netherlands)

<sup>6</sup>Centre for Ecology and Hydrology, Lancaster Environment Center (Lancaster, United Kingdom)

\*Presenting author: Yvan Lebras, email: yvan.lebras@rennes.inra.fr

Euryhalin teleosts can live in freshwater as well as in seawater. The success of their survival to different salinities highly depends on their osmotic acclimation capacities. Objective of the present work is to describe acclimation processes to hypersaline environment in the rainbow trout using functional genomic and genetic approaches.

The first part of this project was the genetic control characterisation of processes linked to seawater acclimation in rainbow trout. Analysed traits were post seawater transfer plasma sodium and chloride levels, and gill weight. Plasma ionic contents were obtained 24h after each of two successive seawater challenges when gill weight was measured only after the second challenge. Thus, unitrait and multitrait QTL analyses allowed us to reveal 14 QTL.

A complementary approach of eQTL detection then allowed, based on a previous gill differential transcriptomic study and on the results of the QTL approach, to offer 69 exclusive candidate genes among them 7 are of prime interest. While most of these genes are proposed as functional candidates, some are suggested positional. Research of predictive gene expression profiles for the plasmatic traits, enhance the relative importance of several genes as Somatotropin-2 and CD40 and especially ES1, a gene among the 7 previously mentioned.

The existence of a genetic component having an influence on the capacities of osmoregulation in teleost is shown here. These results lay the road with cobblestones for a definite investigation of the genetic bases of seawater acclimation processes in teleosts.

# **A comparison of regression models for testing QTL'eQTL co location**

Xiaoqiang Wang<sup>1\*</sup>, Jean Michel Elsen<sup>2</sup>, Olivier Filangi<sup>1</sup> and Pascale Le Roy<sup>1</sup>

<sup>1</sup>INRA, UMR598, Génétique Animale, IFR140 GFAS, 35000 Rennes, France

<sup>2</sup>INRA, UMR 1313, 78352 Jouy-en-Josas, France

\*Presenting author: Xiaoqiang Wang, email: xiaoqiang.wang@rennes.inra.fr

Exploring the genetic architecture in biology relies on the accurate identification of quantitative trait loci (QTL) and gene expression quantitative loci (eQTL). Currently, as more and more eQTL have been mapped in various species, the importance of searching the eQTL which co locate with a given QTL is highlighted. A number of statistical approaches are proposed to look for the eQTL of interest, in particular test whether a same QTL affects two different traits or two different linked QTL affect separately two traits. All the methods were useful and widely applied in eQTL studies. However, the required conditions when using these methods hinder their applications for all the circumstances. This study firstly discusses the suitable conditions for the methods existing in the literature. Then we propose two methods, making use of the trait and QTL information, in order to look for the eQTL co located with the QTL. In addition, the parameters influencing the power for above eQTL mapping are discussed. Finally, through simulated data in an outbreed population, we verify the required conditions for the existed method and validate our methods in co location studies.

## **A simulation-based approach for controlling the false positive rate in genome-wide association studies.**

Marcin Kierczak - Carlborg's Lab, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, SWEDEN

Xia Shen - Carlborg's Lab, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, SWEDEN

Orjan Carlborg - Carlborg's Lab, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, SWEDEN

Precise control of false positive rate (FPR) is an important issue in genome-wide association studies. The most common approach to setting significance thresholds in GWAS experiments is based on Bonferroni-corrected p-values. While efficient in minimizing FPR, this approach is often considered too conservative. Other popular approaches are, e.g.: an optimized FDR approach (q-values) or permutation tests. Each of these has its advantages and limitations.

Here, we propose a simulation-based approach to assess the true FPR in genome-wide association studies. First, we use original data to construct a large number of datasets, each containing an artificial signal at randomly selected markers. Next, for each such dataset, we perform an association test and record the obtained p-values. In each dataset, we consider the artificially introduced signal as a true positive. Thus, we can determine FPR at different significance thresholds. The results averaged over a large number of experiments give stable and reliable measure of FPR at different significance thresholds.

We show how our method performs in comparison to other approaches for controlling FPR and discuss the its possible applications. The method is available as an R-package.

## Mapping QTL from QTL-MAS workshop data using four methods

Naveen Kumar Kadri\*, Golam Reza Dashab, Mohammad Mahdi Shariati, Goutam Sahana

<sup>1</sup>Department of Genetics and Biotechnology, Aarhus University

\*Presenting author: Naveen Kumar Kadri, email: Naveenk.kadri@agrsci.dk

The dataset simulated for the QTL-MAS 15 workshop was analyzed with four different methods. The methods were 1) Single marker regression in a mixed model 2) Haplotype based regression 3) GENMIX (**gene**ology based **mixed** model) and 4) Bayesian variable selection.

In the first method single SNPs were successively fitted in a mixed model where the polygenic effect was fitted as random effect. The significance of the marker effect was then assessed after Bonferroni correction. A similar approach was used in the second method where the marker data was phased first, and haplotypes with four SNPs each were fitted as random effects. In the third method GENMIX, local genealogies were used in a mixed model. Local genealogies are genealogies for the longest chromosome region around a marker that do not require recurrent mutation or recombination. This method allows flexible modeling where both family and population relationships can be included in the model. In the fourth method, Bayesian variable selection, all SNPs were fitted simultaneously in a model with a mixture prior distribution for SNP effects. Most of the SNPs (98%) were assumed to have very small effect on the phenotype and few markers explained most of the genetic variance, a priori. The allocation of each SNP to either of these two distributions was done via Gibbs sampling.

The results from the 4 methods employed were comparable. By employing methods 1 to 4 respectively, 8, 12, 12 and 9 QTLs were detected. In all methods, the QTL with the largest effect on the trait was detected on Chromosome 1. The methods can be compared for their power and false discovery rates when the simulated QTLs are known.

## Bayesian Variable Selection to identify QTL affecting a simulated quantitative trait

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

<sup>1</sup>Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands.

<sup>2</sup>Aarhus University, DJF Department of Genetics and Biotechnology, PO Box 50, 8830 Tjele, Denmark.

<sup>3</sup>Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands; Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, PO Box 80163, 3508 TD Utrecht, The Netherlands.

\*Presenting author: Anouk Schurink, email: anoukschurink@hotmail.com

**Background.** Recent developments in genetic technology and methodology enable accurate detection of QTL and estimation of breeding values, even in individuals without phenotypes. The QTL-MAS workshop offers the opportunity to test different methods to perform a genome-wide association study on simulated data with a QTL structure that is unknown beforehand. The simulated data contained 3,220 individuals: 20 sires and 200 dams with 3,000 offspring. All individuals were genotyped, though only 2,000 offspring were phenotyped for a quantitative trait. QTL affecting the simulated quantitative trait were identified and breeding values of individuals without phenotypes were estimated using Bayesian Variable Selection, a multi-locus SNP model in association studies. The probability of a SNP being modelled in the second distribution ( $\pi$ ) was estimated.

**Results.** Estimated heritability of the simulated quantitative trait was 0.31. Mean posterior probability of SNP modelled in the second distribution, i.e. SNP with large effect, was 0.0075 (95% highest posterior density region: 0.0014-0.0139). The genome-wide association analysis resulted in 18 significant SNP, comprising 6 QTL on chromosome 1, 2 and 3. In total 23 putative SNP, comprising 8 putative QTL were detected.

**Conclusions.** Bayesian Variable Selection using thousands of SNP was successfully applied to genome-wide association analysis of a simulated dataset with unknown QTL structure. Strong and putative QTL were detected and breeding values were estimated for individuals with and without phenotypes.

# Estimation of breeding values and detection of QTL in QTL-MAS 2011 dataset using GBLUP and Bayesian approaches

Marcin Pszczola<sup>123\*</sup>, Anna Wolc<sup>14</sup>, Sebastian Mucha<sup>1</sup>, Maciej Szydlowski<sup>1</sup>, Marek A. Wietrzykowski<sup>1</sup>, Alicja Borowska<sup>1</sup>, Tomasz Strabel<sup>1</sup>

<sup>1</sup> Department of Genetics and Animal Breeding, Poznan University of Life Sciences, Poznan, Poland

<sup>2</sup> Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Lelystad, The Netherlands

<sup>3</sup> Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands

<sup>4</sup> Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, USA

\*Presenting author: Marcin Pszczola, email: marcin.pszczola@gmail.com

**Background.** The goal of this study was to compare estimated breeding values and detected QTL positions obtained using GBLUP and BayesB. Data provided for the 15<sup>th</sup> QTL-MAS Workshop were analyzed. Breeding values for young individuals without phenotypic information were estimated using pedigree BLUP, GBLUP and BayesB methods. Two different marker based relationship matrices were used in GBLUP. Variance components were estimated using REML for each of BLUP procedures separately. QTL were detected with BayesB method and from estimated allele substitution effects from GBLUP analyses. The presence of dominance and epistasis were tested in two separate analyses using pedigree BLUP and GBLUP.

**Results.** Heritability estimates ranged from 0.27 to 0.30. Correlations between breeding values estimated based on marker information were high and ranged from 0.92 to 0.99. Correlations between these breeding values and pedigree BLUP estimates were smaller and ranged from 0.67 to 0.77. GBLUP and BayesB indicated QTL with large effect at the beginning of chromosomes 1 and 3, and at the end of chromosome 2. Several QTL with smaller effects were also detected. Estimated dominance variance was 0.51 compared to the additive variance of 22.19 when GBLUP was applied, whereas the estimated dominance variance was not present when pedigree relationship matrix was used. None of the models detected additive-by-additive epistasis.

**Conclusions.** When a trait has mainly genetic additive background the breeding values estimated by GBLUP are very similar to those obtained with BayesB. Moreover, GBLUP and BayesB map QTL to similar chromosomal regions.

# Genomic Breeding Value Prediction and QTL Mapping of QTLMAS2011 Data Using Bayesian Methods

Jian Zeng<sup>1</sup>, Anna Wolc<sup>1</sup>, Rohan L. Fernando<sup>1</sup>, Dorian J. Garrick<sup>1\*</sup>, and Jack C. M. Dekkers<sup>1</sup>

<sup>1</sup>Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, Iowa 50011, U. S. A.

\*Presenting author: Dorian J. Garrick, email: dorian@iastate.edu

Bayesian methods have been widely used in animal breeding and quantitative trait loci (QTL) mapping to simultaneously evaluate a large number of single nucleotide polymorphisms (SNPs) across the whole genome. Here we present the results of applying the BayesC $\pi$  method to predict genomic breeding values (GEBVs) and map QTL positions using the dataset simulated for the 15th MAS-QTL workshop. Common variance is assumed for all SNP effects fitted in the model in BayesC $\pi$  where  $\pi$ , the proportion of SNPs with zero effect on the trait, is estimated from the data. A variety of models, including only additive covariates, or additive and dominance covariates, with or without polygenic effects estimated from pedigree, were fitted to the data using the BayesC $\pi$  framework. Cross-validation was used for model assessment by separating observations into two equal-sized groups. GEBVs were predicted as the sum of the SNP effects and covariate values derived from the individual genotypes. We mapped QTL to positions where the variance of genomic predictions based only on covariates from windows of 5 consecutive SNPs were markedly higher than the background window variance. The effects of QTL were estimated by the sum of the effects of neighboring SNPs with posterior probability of being fitted in the model larger than 0.1. Results show that dominance genetic variance was trivial suggesting only additive gene-action affects this trait. Compared with an additive model, the model fitting both additive and dominance covariates gave slightly higher prediction accuracy in cross-validation but the superiority is too small to be significant. Including polygenic effect was not influential on predictive accuracy in these data with only two generations of pedigree. A total of nine QTL were identified, with three mapping to chromosome 1, four to chromosome 2 and the remaining two on chromosome 3. Weak signals of association in terms of variance of SNP windows were ignored to reduce the probability of reporting false positives.

# Variance Controlling Genes: Significant Contributors to The Missing Genetic Variation for Complex Traits

Xia Shen<sup>1\*</sup>, Mats Pettersson<sup>2</sup>, Lars Rönnegård<sup>3</sup>, Örjan Carlborg<sup>4</sup>

<sup>1</sup>Department of Cell and Molecular Biology, Uppsala University, Uppsala Sweden. Statistics Unit, Dalarna University, Borlänge, Sweden.

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

<sup>3</sup>Statistics Unit, Dalarna University, Borlänge, Sweden. Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.

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

\*Presenting author: Xia Shen, email: xia.shen@lcb.uu.se

One of the current analytical challenges in genetics is how to process data and interpret results from genome-wide association studies (GWAS). Using a publicly available GWA dataset from *Arabidopsis thaliana*, we show that mapping of loci controlling the variance, rather than the mean, identifies both individual loci with strong effects as well as interacting loci that act as regulators in gene-by-gene and gene-by-environment interactions. The variance-controlling loci make significant contributions to the missing genetic variation for many complex traits, where a significant portion of the environmental variance in a standard GWAS is actually heritable. The previously undetected loci explain large fractions of the phenotypic variance - up to 30% as individual factors and up to 60% through interactions. By identifying these loci, we provide new insights to the genetic architecture of the studied traits, illustrate the benefit of using alternative genetic models for analyzing GWAS data and suggest that a paradigm shift in the genetic modeling of complex traits is needed to properly study the total contribution of genetics to multifactorial trait expression.

Shen, X., Pettersson, M., Rönnegård, L. and Carlborg, Ö. (2011). ***Inheritance beyond plain heritability: variance-controlling genes in Arabidopsis thaliana.*** (Submitted)



# Use of planes in genotype space to detect high order epistasis

Mats Pettersson<sup>1\*</sup>

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

\*Presenting author: Mats Pettersson, email: mats.pettersson@slu.se

High order epistasis is difficult to detect, since it typically requires both a large data set and a large number of statistical tests. In order to make detection feasible, we propose the use of “planes” or “slices” from a three-dimensional genotype-phenotype (GP) map to define a computationally efficient measure based on the difference in variance between these planes. For a triplet of loci (say Q1, Q2 and Q3) that affects a given trait, the GP map can be split into three planes that each contains all three-locus genotypes that share a genotype at one stratifying locus. That is, if the stratifying locus is Q1, the planes will consist of the genotypes that include 11, 12, and 22 at this locus respectively. If Q1 does not interact with either Q2 or Q3, the planes will have the same shape and thus the same variance. By conditioning on all three loci, it is possible to find any interaction pattern within this triplet, while filtering out the signal from marginal effects. Since variance is computationally cheap to calculate, this approach is much more feasible to apply to a large data set – which remains a prerequisite – than a more complex scoring method, such as regression.

To illustrate the concept, I describe results based on data from the “Virginia lines” of divergently selected chicken; a system where epistatic interactions have been previously shown to be of importance to the selected trait (body weight at 8 weeks of age; Carlborg et al, 2006).

*Carlborg O, Jacobsson L, Ahgren P, Siegel P, Andersson L: (2006) Epistasis and the release of genetic variation during long-term selection. Nat Genet. 38:418-420.*

# A problem of locating multiple interacting QTL by logic regression

Magdalena Malina<sup>1\*</sup>, Małgorzata Bogdan<sup>2</sup>

<sup>1</sup>Mathematical Institute, University of Wrocław

<sup>2</sup>Institute of Mathematics and Computer Science, Wrocław University of Technology

\*Presenting author: Magdalena Malina, email: malina@math.uni.wroc.pl

**Background.** Logic regression ([2]) is an extension of the classical regression model, designed for the situation where the explanatory variables are binary. In logic regression, the predictors are defined as logic expressions operating on these binary variables. Within this framework, the interaction is usually understood in the colloquial way - it is simply a “causal” logic expression, which directly describes the dependence of the response variable on the specific combination of values of the explanatory variables. Since this definition of interaction agrees with its common understanding, logic regression models are often easier to interpret than the standard linear models.

**Results.** Considering the problem of detection of interacting QTL as a problem of multiple testing, we show theoretically that logic regression may lead to a larger power of detection an interaction than the standard methods based on a Cockerham's genetic model. We applied a method based on the logic regression - LogicFS ([1]) - for QTL mapping and compared it to the classical methods based on the standard linear model and the standard methodology implemented in R/QTL. We considered the two data sets: <http://phenome.jax.org/phenome/protodocs/QTL/QTL-Lyons3.xls> and <http://churchill.jax.org/datasets/ctlarchive/fatpads.shtml>. The first one contains different phenotypes related to cholesterol gallstones formation in an intercross population of inbred mice. The considered phenotype is binary (a presence of gallstones). The second dataset considers phenotypes related to obesity in an intercross population of inbred mice. The considered phenotype is a total gonadal fat pads weight.

**Conclusions.** Our results show that LogicFS can be successfully applied to identification of interacting QTL. Specifically for both of the analyzed data sets, LogicFS was able to detect some interactions which are not reported in original articles and are not detected by standard QTL mapping methodology.

[1] Schwender, H. and Ickstadt, K. (2008). Identification of SNP interactions using logic regression. *Biostatistics*, 9. 187-198

[2] Ruczinski I., Kooperberg C., LeBlanc M., *Logic regression*, *J. Comput. Graphical Statist.* 12 (3), (2003), 474-511

# qtl.outbred: Interfacing outbred line cross data with the R'qtl mapping software

Ronald M. Nelson<sup>1\*</sup>, Xia Shen<sup>2</sup>, Örjan Carlborg<sup>1,2</sup>

<sup>1</sup>Department of Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, SE-75007 Uppsala, Sweden.

<sup>2</sup>Department of Cell and Molecular Biology, Uppsala University, BMC Box 598, SE-75124 Uppsala, Sweden. Ronald M. Nelson<sup>1</sup>

\*Presenting author: Ronald Nelson, email: ronnie.nelson@slu.se

Outbred line cross data provide an additional challenge compared to inbred line cross data. This is due to the uncertainty in the origin of the alleles in the offspring since they inherit their genes from non-homozygous parent populations. Software developed to analyse inbred line cross data are therefore unsuitable when outbred line cross data are available. We developed **qtl.outbred** as an umbrella package that calculate genotype probabilities from outbred line cross data and import it into the user friendly software package R/qtl (Broman, 2003). Using **qtl.outbred**, the genotype probabilities from outbred line cross data are calculated by interfacing with a new and efficient algorithm (Crooks, in press) developed for analyzing arbitrarily large datasets (included in the package) or imported from other sources such as the web-based tool, GridQTL (Seaton, 2006). This package enables the user to analyse outbred line cross data with the same ease and speed as inbred line cross data. **qtl.outbred** is run in the statistical environment, R.

*Broman KW. R/qtl: QTL mapping in experimental crosses. Bioinformatics 2003, 19:889-890*

*Crooks L, Nettelblad C, Carlborg Ö. An improved method for estimating chromosomal line origin in QTL analysis of crosses between outbred lines. G3: Genes, Genomes, Genetics (in press)*

*Seaton G, Hernandez J, Grunchev JA, White I, Allen J, de Koning DJ, Wei W, Berry D, Halley C, Knott SA. GridQTL: a grid portal for QTL mapping of compute intensive datasets. In Proceedings of 8th World Congress on Genetics Applied to Livestock Production. Brasil 2006*

## **AnnotQTL: a new tool to gather functional and comparative information on a genomic region**

Lecerf F.<sup>1,2</sup>, Bretaudeau A.<sup>3</sup>, Sallou O.<sup>3</sup>, Desert C.<sup>1,2</sup>, Blum Y.<sup>1,2,4</sup> Lagarrigue S.<sup>1,2</sup> and Demeure O.<sup>1,2\*</sup>

<sup>1</sup>INRA, UMR598 Génétique Animale, F-35000 Rennes, France.

<sup>2</sup>Agrocampus OUEST, UMR598 Génétique Animale, F-35000 Rennes, France.

<sup>3</sup>GenOuest Platform, INRIA/Irisa – Campus de Beaulieu, F-35042 Rennes Cedex, France.

<sup>4</sup>Agrocampus OUEST, Applied Mathematics Department, Rennes, France.

\*Presenting author: Olivier Demeure, email: [olivier.demeure@rennes.inra.fr](mailto:olivier.demeure@rennes.inra.fr)

AnnotQTL is a web tool designed to aggregate functional annotations from different prominent websites by minimizing the redundancy of information. Although thousands of QTL regions have been identified in livestock species, most of them are large and contain many genes. This tool was therefore designed to assist the characterization of genes in a QTL interval region as a step towards selecting the best candidate genes. It localizes the gene to a specific region (using NCBI and Ensembl data) and adds the functional annotations available from other databases (Gene Ontology, Mammalian Phenotype, HGNC and Pubmed). Both human genome and mouse genome can be aligned with the studied region to detect synteny and segment conservation, which is useful for running inter-species comparisons of QTL locations. Finally, custom marker lists can be included in the results display to select the genes that are closest to your most significant markers. We use examples to demonstrate that in just a couple of hours, AnnotQTL is able to identify all the genes located in regions identified by a full genome scan, with some highlighted based on both location and function, thus considerably increasing the chances of finding good candidate genes. AnnotQTL is available at <http://annotqtl.genouest.org>.

## **GPU accelerated QTLMap**

Guillaume Chapuis<sup>1\*</sup>, Jean Michel Elsen<sup>2</sup>, Olivier Filangi<sup>3</sup>, Dominique Lavenier<sup>1</sup>, Pascale Le Roy<sup>3</sup>

<sup>1</sup>IRISA - Symbiose Campus universitaire de Beaulieu, 35042 Rennes Cedex, France

<sup>2</sup>INRA, UMR 1313, 78352 Jouy-en-Josas, France

<sup>3</sup>INRA, UMR598 Génétique Animale, F-35000 Rennes, France.

\*Presenting author: Guillaume Chapuis, email: guillaume.chapuis@irisa.fr

Recent years have seen a stagnation of processors clock speed. Accelerating execution times of programs nowadays can no longer be achieved by simply upgrading one's hardware. The difference with today's computing devices is that they can handle an increasing number of tasks simultaneously, leaving to the programmers the responsibility of extracting parallelism from algorithms. For a program to be parallelised, its algorithm needs to be composed almost entirely of independent tasks. This is the case for QTL analysis ; moreover QTL analysis as implemented in QTLMap exhibits a great amount of embarrassingly parallel tasks, making it an ideal candidate for execution on Graphics Processing Units.

Graphics cards architecture allows simultaneous execution of a single program over multiple independent data. Though much slower than a CPU core in terms of clock rate, GPU multiprocessors draw their power from executing up to a thousand similar tasks at the same time. Almost no synchronization can be performed between simultaneous tasks on a GPU, therefore GPU computing can only be used for a very small subset of algorithms.

Our new implementation of QTLMap takes full advantage of both a GPU and a multicore central processor. Embarrassingly parallel tasks, such as matrix triangulation, are computed on the GPU while more dependent tasks are executed on several CPU cores. Speedups are achieved by overlapping as much as possible these executions. Our GPU/multicore implementation performs up to 20 times faster than the previous multicore implementation and allows extensive QTL analysis to be conducted in a reasonable time, while maintaining the same level of precision. When double precision is not mandatory, our implementation of QTLMap can also be used in single precision, yielding to even greater speedups.

## **Strategies to improve the accuracy of genomic predictions**

Dorian J. Garrick<sup>1\*</sup>

<sup>1</sup>Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, Iowa 50011, U. S. A.

\*Presenting author: Dorian J. Garrick, email: [dorian@iastate.edu](mailto:dorian@iastate.edu)

The intent of genomic prediction is to identify markers in strong linkage disequilibrium with QTL in order to rank selection candidates at an early age. It was first thought that whole genome methods that simultaneously fit many markers relied entirely on linkage disequilibrium to obtain predictive power. However, it turns out that these methods are also very good at exploiting linkage. This talk will review the role of linkage disequilibrium in genomic prediction and indicate some of the consequences of predictions that jointly exploit linkage and linkage disequilibrium.

## Could genomic selection methods be efficient to detect QTL?

C. Colombani<sup>1\*</sup>, P. Croiseau<sup>2</sup>, C. Hozé<sup>3</sup>, S. Fritz<sup>3</sup>, F. Guillaume<sup>4</sup>, D. Boichard<sup>2</sup>, A. Legarra<sup>1</sup>, V. Ducrocq<sup>2</sup>, C. Robert-Granié<sup>1</sup>

<sup>1</sup>INRA, UR631-SAGA, BP 52627, 31326 Castanet-Tolosan Cedex, France.

<sup>2</sup>INRA, UMR1313-GABI, 78352 Jouy-en-Josas, France.

<sup>3</sup>UNCEIA, 149 rue de Bercy, 75595 Paris, France, <sup>4</sup>Institut de l'Élevage, 149 rue de Bercy, 75595 Paris, France.

\*Presenting author: Carine Colombani, email: carine.colombani@toulouse.inra.fr

With the discovery of thousands of single nucleotide polymorphisms (SNPs), QTL associated with traits of interest can be localised with higher precision. Linkage Disequilibrium and Linkage Analysis (LDLA) approaches are used to detect QTL when the population structure is appropriate. Since June 2010, the French genomic selection (GS) in dairy cattle relies on a Fernando-Grossman model where the list of QTL included in the model comes from two different sources, either an LDLA or a variable selection approach. The GS accuracies were quite similar between the two procedures but the selected SNPs were not strictly the same. This study aims at exploring the differences to assess whether the variable selection approaches could be considered as QTL detection methods.

We tested two strategies of detection of relevant SNPs based on three genomic selection methods: Elastic Net (EN), PLS and Sparse PLS. The PLS regression is a dimension reduction method which projects the independent variables onto latent structures. Sparse PLS adds a penalization step to PLS. EN is a linear combination of Ridge Regression and LASSO estimators. These three methods require the choice of parameters which influence the number of SNP retained in the model. The first strategy was based on applying the genomic selection methods to find the best prediction equation with a maximum of 10, 100 or 500 SNPs. An alternative strategy consisted in running the methods with the optimum parameters in a genomic selection context to predict the SNP effects. Then, the 10, 100 and 500 SNPs with the largest estimated effects were selected. The different SNP sets from the French Holstein breed of 2976 bulls and three traits (Protein Yield, Fat Content and Conception Rate) were compared to LDLA results based on distances between markers.

Results indicated that genomic selection methods were able to locate SNPs near the largest QTL defined by LDLA but also other genome regions undetected by LDLA. In contrast with LDLA which is a uni-QTL approach, GS methods estimates SNP effects simultaneously, this may lead to improve estimates. Since different GS methods highlight different SNPs, further investigations for example based on simulations are required to identify the optimal approach.

# Linear models for breeding values prediction in haplotype-assisted selection

Anna Mucha<sup>1\*</sup>, Heliodor Wierzbicki<sup>1</sup>

<sup>1</sup>Wrocław University of Environmental and Life Sciences, Department of Genetics and Animal Breeding, Kozuchowska 7, 51-631 Wrocław, Poland

\*Presenting author: Anna Mucha, email: anna.mucha@up.wroc.pl

**Background.** The aim of this study was to estimate haplotype effects and then to predict breeding values using linear models. The haplotype based analysis enables avoidance of losing information due to linkage disequilibrium between single markers. There are also less explanatory variables in the linear model which make the estimation more reliable.

**Methods.** Different methods and criteria for marker and haplotype selection were considered. First, markers with MAF lower than 0.01 were excluded from the data set. Then, SNPs in complete linkage disequilibrium were selected. Next step was to construct haplotypes and to estimate their frequencies basing on selected SNPs. The haplotypes with a frequency lower than 1% were not considered in further analysis. Chosen haplotypes were used as the explanatory variables in the linear model for breeding values prediction. Linear models with fixed and random haplotype effects were tested.

**Results and Conclusions.** 3234 out of 9990 markers failed MAF criterion. The haplotypes were constructed basing on different numbers of SNPs for each chromosome: 211 SNPs in 92 subsets for the first chromosome, 178 SNPs in 87 subsets for the second chromosome, 150 SNPs in 65 subsets for the third chromosome, 166 SNPs in 73 subsets for the fourth chromosome and 216 SNPs in 92 subsets for the fifth chromosome. Haplotypes effects were expressed as the deviation from the average phenotypic value in the population.

Hayes B., Hagesæther N., Ådnøy T., Pellerud G., Berg P.R., Lien S. (2006): Effects on Production Traits of Haplotypes Among Casein Genes in Norwegian Goats and Evidence for a Site of Preferential Recombination. *Genetics* 174: 455-464 (September 2006)

Kolbehdari D., Schaeffer L.R., Robinson J.A.B. (2007): Estimation of genome wide haplotype effects in half-sib designs. *Journal of Animal Breeding and Genetics* 2007 Dec; 124(6):356-61

Sham P.C., Rijdsdijk F.V., Knight J., Makoff A., North B., Curtis D. (2004): Haplotype Association Analysis of Discrete and Continuous Traits Using Mixture of Regression Models. *Behavior Genetics*, vol. 34, No. 2, March 2004.



## Effect of the prior distribution of SNP effects on the estimation of total breeding value

Nadaf J<sup>1</sup>, Riggio V<sup>2</sup>, Yu T-P<sup>1</sup>, Pong-Wong R<sup>2</sup>

<sup>1</sup>DNA Landmarks Inc. St-Jean-Sur-Richelle, J3B 6X3 Quebec, Canada.

<sup>2</sup>The Roslin Institute and the R(D)SVS, The University of Edinburgh, Easter Bush, EH25 9RG, UK.

Five main methods commonly applied in genomic selection were used to calculate the GEBV on the 15th QTLMAS workshop dataset: GBLUP, LASSO, Bayes A and two Bayes B type of methods (Bayes Bn and Bayes Bt). GBLUP is a mixed model approach where GEBVs are obtained using a relationship matrix calculated from the SNP genotypes. The remaining are regression-based approaches where the SNP effects are estimated and, then GEBVs are calculated given the individuals' genotypes.

The differences between the regression-based methods are on their prior distributions for the SNP effects. The prior distribution for LASSO is a Laplace distribution, for Bayes A is a scaled Student-t distribution, and the Bayes B type methods have a Spike and Slab prior where only a proportion ( $\pi$ ) of SNP has an effect, following a given distribution. In this study, two different distributions were considered for the BB methods: (i) normal and (ii) scaled Student-t. They are referred here as the Bayes Bn and Bayes Bt methods, respectively.

These prior distributions are defined by one or more parameters controlling their scale ( $\lambda$ ), shape (df) or proportion of SNP with effect ( $\pi$ ): LASSO requires one ( $\lambda$ ); two for Bayes A ( $\lambda$ , df) and Bayes Bn ( $\lambda$ ,  $\pi$ ); and three for Bayes Bt ( $\lambda$ , df,  $\pi$ ). In this study, all parameters were estimated from the data. An extra scenario for Bayes A and Bayes Bt was included where df was not estimated but fixed to 4. For GBLUP, the heritability was also estimated from the data. The implementation of GBLUP was done using ASREML and the remaining methods were implemented using a MCMC approach.

Generally speaking, the GEBV with the different methods were consistent with the correlation between the different estimates ranging from 0.898 to 1. When estimating all parameters, the results for Bayes A and Bayes Bt were the same ( $r=1$ ). The results for Bayes Bn was similar to Bayes Bt with the degree of freedom fixed to 4 ( $r=0.999$ ). The results of LASSO and GBLUP were similar to each other ( $r=0.991$ ) and slightly less correlated with the other methods.

# Approximate Bayesian prediction of genomic breeding values and variances using INLA

Patrik Waldmann<sup>1</sup> and Johann Sölkner<sup>1</sup>

<sup>1</sup>Division of Livestock Sciences, Department of Sustainable Agricultural Systems, University of Natural Resources and Applied Life Sciences, Gregor Mendel Str.33, A-1180, Vienna, Austria

\*Presenting author: Patrik Waldmann, email: patrik.waldmann@boku.ac.at

Bayesian inference of genomic breeding values (GEBV) and variances has become a subject of major concern. MCMC methods have been very successful and played a key role in Bayesian analysis, but these methods tend to be computationally demanding in large scale high dimensional problems. Hence, it has been suggested that approximate Bayesian inference could be a feasible alternative to MCMC. One recently introduced Bayesian method is integrated nested Laplace approximation (INLA). INLA substitutes MCMC simulations with accurate, deterministic approximations to posterior marginal distributions. Within the INLA framework, the GEBVs are modelled as latent Gaussian variables over a structured precision matrix (the inverse of the genomic relationship matrix). Based on analysis of the 3220 individuals in the simulated QTLMAS 2011 data, we show that INLA can predict the full posterior distributions of the GEBVs for individuals with and without phenotypic observations in less than an hour. The estimates of the mean genetic and error variances were 21.9 (95% CI: 16.8 – 29.1) and 59.3 (95 % CI: 55.4 – 63.6), respectively.

## Genomic selection: implementation in dairy cattle

V. Ducrocq<sup>1\*</sup>, S. Fritz<sup>2</sup>, F. Guillaume<sup>1,3</sup>, P. Croiseau<sup>1</sup>, J.J. Colleau<sup>1</sup> D. Boichard<sup>1</sup>

<sup>1</sup>INRA, UMR1313-GABI, 78352 Jouy-en-Josas, France, <sup>2</sup>UNCEIA, 149 rue de Bercy, 75595 Paris, France, <sup>3</sup>Institut de l'Élevage, 149 rue de Bercy, 75595 Paris, France.

\*Presenting author: Vincent Ducrocq, email: vincent.ducrocq@jouy.inra.fr

The possibility to precisely estimate breeding values at birth using genotype information has been presented as a major breakthrough which will boost genetic gain in dairy cattle breeding schemes. A wide panel of genomic evaluation methods have been proposed. But in fact, a much larger variability exists worldwide on how genomic selection (GS) is being implemented in practice.

A major source of heterogeneity is the breed: population size, geographical extension (international / national / local), and opportunities for international collaborations have a tremendous impact on the size of reference populations and therefore on GS accuracy. This accuracy is affected not only by trait heritability but also by extent of performance recording and existence of preferential treatment. First implementations were based on AI bull reference populations but using own performances will be necessary in the future to update these reference populations and in populations or for traits without large scale progeny test.

GS should logically result in a huge reduction of generation interval on the male side, and consequently, the ending of the lengthy and costly progeny test. In fact, not all breeding companies seem to be prepared to go in this direction. Reasons for this vary from the necessity to provide evidence that GS is working to reluctance to give up the “star system” where a few top bulls with high accuracy proofs are marketed internationally. GS is then mainly used for pre-selection before progeny testing. The impact of such an evolution on genetic variability may be quite detrimental. Simulations show large increases in expected genetic gain (+50 to 100%) in nearly all implementations without progeny test. Using many bulls as sires of bulls and cows leads to the lowest increase in inbreeding (-20% in comparison with the progeny test scenario).

On the female population, no reduction in generation interval can be expected unless sexed semen is heavily used. But the much higher accuracy of selection on low heritability traits (e.g., functional traits) could have a significant impact on future genetic gains and on selection objectives. Again, the current number and origin of females genotyped greatly fluctuate between countries.

Other key aspects are playing or will play a role in the future on how GS is implemented. These include the development of efficient multi-breed and/or international genomic evaluations, the use of information from sequenced key animals, the collection of new phenotypes (e.g., on health traits or fine milk composition) or the national and international restructuring of the breeding industry with possibly new players entering.

In contrast with what is often thought, the future efficiency of breeding schemes will be more a function on how GS is implemented than on the size of reference populations.

# **A simple two step Bayesian approach for genomic prediction of breeding values**

M. Mahdi Shariati<sup>1\*</sup>, Peter Sorensen<sup>1</sup>, Luc Janss<sup>1</sup>

<sup>1</sup>Department of Genetics and Biotechnology, Aarhus University

\*Presenting author: Mohammad Mahdi Shariati, email: mohammad.shariati@agrsci.dk

Breeding values of animals with marker data were estimated using a two step Bayesian approach. SNP markers were ranked based on their effects and markers with similar estimated effects were grouped together. In step 1, all markers with minor allele frequency less than 0.01 were included in the prediction model. Standard deviations of marker effects were assumed to have a censored normal distribution and conditionally on the marker variance, the marker effect had a normal distribution leading to a tick tailed marginal prior distribution for marker effects. In step 2, markers were ranked based on their estimated variance on the trait in step 1. Each 100 markers were assigned to one group with a common variance. By this, instead of specifying one variance for the conditional distribution of each marker, one variance was specified for 100 markers with similar estimated effects. In a further analysis, only 450 markers with largest effects in step 2 were kept in the prediction model such that each 50 markers were assigned to a group of markers with similar estimated effects. Our results showed that both of these approaches outperformed well known Bayesian models with either a common prior or a mixture prior for marker effects. In conclusion, grouping markers is less flexible than allowing each marker to have a specific marker variance but, by grouping, the power to estimate marker variances increases and more reliable breeding values are achieved.

# Genome-wide association study of the 15th QTL-MAS workshop data using mixed model based single locus regression analysis

Weixuan Fu<sup>1</sup>, Jianfeng Liu<sup>1</sup>, Xiangdong Ding<sup>1</sup>, Chonglong Wang<sup>1,2\*</sup>, Zhe Zhang<sup>1</sup>, Peipei Ma<sup>1</sup>, Ziqing Weng<sup>1</sup>, Qin Zhang<sup>1\*\*</sup>

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

<sup>2</sup>Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei 230031, China

\*Presenting author, email: ahwchl@163.com

\*\*Corresponding author, email: qzhang@cau.edu.cn

**Background.** Mixed model based single locus regression analysis (MMRA, Jiang et al., 2010) was used to analyze the data of 15<sup>th</sup> QTL-MAS 2011 workshop to detect potential significant associated SNPs. A Wald chi-squared statistic with  $df=1$  was constructed to examine whether the SNP was associated with the simulated trait. Moreover, phenotypic observations were permuted 10,000 times against the genotype and pedigree data, to obtain significance thresholds for the associated SNPs. And linkage disequilibrium (LD) between SNPs for all the animals was quantified as  $r^2$  and the LD block was defined by the criteria of Gabriel et al. (2002), to indicate QTL regions.

**Result.** Applying an animal model in DMUv6 showed the heritability of the quantitative trait was approximately 57%. With the genome-wide association analysis using MMRA, we detected 104 significant SNPs ( $P < 0.05$ ) for the trait by simple Bonferroni correction; with permutation test we reduced the list to 59 SNPs. Through the definition of the LD blocks, our analysis confirmed 15 significant putative QTL regions. There were 12, 1 and 2 QTL regions on chromosome 1, 2 and 3, respectively, and none of the QTL appeared on chromosome 4 and 5.

**Conclusion.** MMRA is a fast computing method with feasibility of performing time-consuming permutation test. Using LD block can effectively combine significant SNPs for QTL regions detection.

[1]Jiang L., Liu J., Sun D., Ma P., et al. (2010) Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS One* 5, e13661.

[2]Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M. et al. (2002) The structure of haplotype blocks in the human genome. *Science* 296, 2225-9.

# Genomic breeding value estimation for the common dataset of the 15<sup>th</sup> QTL-MAS Workshop

Chonglong Wang<sup>1,2\*</sup>, Peipei Ma<sup>1</sup>, Zhe Zhang<sup>1</sup>, Xiangdong Ding<sup>1</sup>, Jianfeng Liu<sup>1</sup>, Ziqing Weng<sup>1</sup>, Weixuan Fu<sup>1</sup>, Qin Zhang<sup>1\*\*</sup>

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

<sup>2</sup>Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei 230031, China

\*Presenting author, email: ahwchl@163.com

\*\*Corresponding author, email: qzhang@cau.edu.cn

**Background.** Genomic breeding values estimation is the key to genomic selection. Among many approaches, BLUP methods and Bayesian methods are most commonly used for estimating genomic breeding values. Here, we applied two BLUP methods and three Bayesian methods to the common dataset provided by the 15<sup>th</sup> QTL-MAS Workshop to evaluate and compare their predictive performances.

**Methods.** The variance components and the traditional BLUP estimated breeding values (EBVs) based on trait phenotypes and pedigree were obtained from DMUv6. The genomic estimated breeding values (GEBVs) were obtained by TABLUP (Zhang et al., 2010), GBLUP (VanRaden, 2008), BayesA, BayesB (Meuwissen et al., 2001), and BayesC<sub>π</sub> (Fernando et al.). The MCMC chains for BayesA, BayesB and BayesC<sub>π</sub> were run for 50,000 cycles and the first 5,000 cycles were discarded as burn in.

**Results.** The additive genetic and environmental variance were estimated as 55.79 and 42.50, respectively. For the 20 sires, the reliability of traditional EBVs was 0.95, and the correlations between EBVs and GEBVs by different methods ranged from 0.94 to 0.97. The highest correlation was given by GBLUP and the lowest by BayesC<sub>π</sub>. For the 1000 progeny without phenotypic values, the correlations between GEBVs by different methods ranged from 0.83 to 0.99. The highest correlation was between TABLUP and BayesB, and the lowest between GBLUP and BayesC<sub>π</sub>.

**Conclusions.** The TABLUP method is an improvement of the GBLUP method. Results produced by TABLUP are consistent with those produced by Bayesian methods. TABLUP is therefore a promising method for genomic breeding value estimation in the candidate populations because of its advantages of being easy to obtain reliabilities of GEBVs and easily extended to estimate GEBVs for individuals without genotypes and for multiple traits.

[1] Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.

[2] VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91: 4414–4423.

[3] Zhang Z, Liu J, Ding X, Bijma P, de Koning D-J, et al. (2010) Best Linear Unbiased Prediction of Genomic Breeding Values Using a Trait-Specific Marker-Derived Relationship Matrix. *PLoS ONE* 5(9): e12648.

## Alternative strategies for selecting subsets of predicting SNPs by LASSO-LARS procedure

M. Graziano Usai<sup>1</sup>, Antonello Carta<sup>1</sup>, Sara Casu<sup>1</sup>

<sup>1</sup>*Settore Genetica e Biotecnologie, AGRIS-Sardegna, Olmedo 07040, Italy*

\*Presenting author: Graziano Usai, email: graziano.usai@gmail.com

The least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996) can be used to predict SNP effects. This operator has the desirable feature of including in the model only a subset of explanatory SNPs, which can be useful both in QTL detection and GWS studies. LASSO solutions can be produced by the least angle regression (LARS; Efron et al., 2004) algorithm. The big issue with this procedure is to define the best constraint ( $t$ ), *i.e.* the upper bound of the sum of absolute value of the SNP effects which roughly corresponds to the number of SNPs to be selected. Usai et al. (2009) dealt with this problem by a cross-validation approach and defined  $t$  as the average number of selected SNPs overall replications. Nevertheless, in small size populations, such parameter could give underestimated values of  $t$ .

Here we propose two alternative ways to define  $t$  and compared them with the “classical” one. The first (strategy 1), was based on 1,000 cross-validations carried out by randomly splitting the reference population (2,000 individuals with performance) into two halves. The value of  $t$  was the number of SNPs which occurred in more than 5% of replications. The second (strategy 2), which did not use cross-validations, was based on the minimization of the Cp-type selection criterion which depends on the number of selected SNPs and the expected residual variance (estimated by GBLUP).

The size of the subset of selected SNPs was 46, 189 and 64 for the classical approach, strategy 1 and 2 respectively. Classical and strategy 2 gave similar results and indicated almost the same distinct regions (around 15). Strategy 1 confirmed such regions and added further positions which gave a less clear scenario. Correlation between GEBVs estimated with the three strategies and EBVs estimated by GBLUP in daughters without phenotypes were 0.89, 0.94 and 0.91 for classical, strategy 1 and 2 respectively. This suggests the use of strategy 1 for GWS and the other two for QTL detection.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. (2004). *Annals of Statistics* 32, 407–499.

Tibshirani, R. (1996). *Journal of the Royal Statistical Society. Series B* 58, 267–288.

Usai, M. G., Goddard, M. E., Hayes, B. J. (2009). *Genetics Research* 91, 427–436.

# Genomic selection using regularized linear regression models: ridge regression, lasso and elastic net and their extensions

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

<sup>1</sup>Bioinformatics Unit, Institute for Crop Science, University of Hohenheim, Fruwirthstrasse 23, 70599 Stuttgart, Germany

\*Presenting author: Joseph O Ogutu, email: jogutu2007@gmail.com

## Background

Genomic selection (GS) is emerging as an efficient and cost-effective method for estimating breeding values using molecular markers distributed over the entire genome. In essence, it involves estimating the simultaneous effects of all genes or chromosomal segments and combining the estimates to predict the total genomic breeding value (GEBV). Accurate prediction of GEBVs is a central and recurring challenge in plant and animal breeding. The existence of a bewildering array of approaches for predicting breeding values using markers underscores the importance of identifying approaches able to efficiently and accurately predict breeding values. Here, we comparatively evaluate the predictive performance of six regularized linear regression methods—ridge regression, ridge regression BLUP, lasso, adaptive lasso, elastic net and adaptive elastic net— for predicting GEBV using dense SNP markers.

## Methods

We predicted GEBV for a quantitative trait using a dataset on 3000 progenies of 20 sires and 200 dams and an accompanying genome consisting of five chromosomes with 9990 biallelic SNP-marker loci simulated for the QTL-MAS 2011 workshop. We applied all the six methods that use penalty-based (regularization) shrinkage to handle datasets with far more predictors than observations. The lasso, elastic net and their adaptive extensions further possess the desirable property that they simultaneously select relevant predictive markers and optimally estimate their effects. The regression models were trained with a subset of 2000 phenotyped and genotyped individuals and used to predict GEBVs for the remaining 1000 progenies without phenotypes. A fivefold cross-validation was used to assess the predictive accuracy of each model, measured as the Pearson correlation between GEBVs and the simulated values.

## Results

The two ridge regression models were outperformed by lasso, elastic net and adaptive lasso but outperformed the adaptive elastic net, which had the lowest accuracy ( $r = 0.429$ , range 0.401–0.467). The correlation between the predicted and simulated breeding values from the fivefold cross-validation was highest for the elastic net ( $r = 0.507$ , range 0.449 – 0.531) followed by the lasso ( $r = 0.506$ , range 0.447– 0.529).

## Conclusions

The elastic net and the lasso both predicted GEBV better than their adaptive extensions, ridge regression, or ridge regression BLUP.



## Molecular Marker Application to Incorporate Salinity Tolerance to West Africa Rice Varieties

Isaac Kofi Bimpong<sup>1\*</sup>, Baboucarr Manneh<sup>1</sup>, Glenn B. Gregorio<sup>2</sup>, Abdelbagi M. Ismail<sup>2</sup>, RK Singh<sup>3</sup>, and Moussa Sie<sup>4</sup>

<sup>1</sup>Africa Rice Centre, Sahel station, B.P 96, Saint Louis, Senegal

<sup>2</sup>International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila, Philippines

<sup>3</sup>IRRI-WARDA office, Mikocheni, P. O. Box 33581, Dar-Es-Salaam, Tanzania

<sup>4</sup>Africa Rice Centre, 01 BP 2031, Cotonou, Benin

\*Presenting author: Isaac Kofi Bimpong, email: kofibimpong@yahoo.com

Salinity is a serious constraint for rice production in West Africa. Molecular markers are now routinely being used to identify and transfer useful genes/QTLs for salinity tolerance. Our goal is to precisely transfer QTLs conferring salinity tolerance into popular African rice varieties to enable higher and more stable yields for farmers living off these marginal lands. One of the major QTL (*Saltol*) derived from the salt-tolerant Indian landrace Pokkali, has been mapped on chromosome 1. The map-based cloning of a gene providing salinity tolerance at the seedling stage (*SKC1*) in a similar position as *Saltol* in the cultivar Nona Bokra was previously published. The QTL for salt tolerance and *SKC1*, maintains K<sup>+</sup> homeostasis in the salt-tolerant cultivars under salt stress, and the *SKC1* gene encodes a member of HKT-type transporters. The Africa Rice Centre aims to employ a marker-assisted backcrossing (MABC) approach to introgress *Saltol* into several African varieties that will subsequently be deployed for commercial cultivation.

Fine-mapping of the *Saltol* QTL and the development of *indel* markers based on candidate genes in this region are being pursued at IRRI. MABC has been used to transfer the Pokkali *Saltol* QTL into popular varieties such as IR64, BR28, and BR11. A diverse set of germplasm including the upland and lowland NERICAs (New Rice for Africa), collection from Cassamance regions in Senegal and Gambia are also being characterized for physiological traits underlying salinity tolerance to identify new sources of tolerance for use as donors in breeding and for identification of novel QTLs. Besides, MABC is currently being used to introgress *Saltol* into few varieties popular in West Africa. The long term goal is to identify and combine genes and QTLs controlling different physiological mechanisms to achieve a higher level of salt tolerance in high yielding West African rice varieties.

# Genome-wide association study for longevity in Fleckvieh cattle

Patrik Waldmann<sup>1</sup>, Chontida Phuthaworn<sup>1</sup>, and Johann Sölkner<sup>1</sup>

<sup>1</sup>Division of Livestock Sciences, Department of Sustainable Agricultural Systems, University of Natural Resources and Life Sciences, Gregor Mendel Straße 33, A-1180 Vienna, Austria

**Background.** Longevity has for some years been of major interest in cattle breeding programs in Austria in order to prolong the productive herd life and thereby reduce the cost of replacements. The objective of this study was to find chromosome regions contributing to the genetic variation of longevity with single SNP and multiple SNP analyses. The phenotypes (daughter yield deviations) for functional herd life of 1935 Fleckvieh bulls were regressed against 41008 SNPs distributed across the whole genome.

**Result.** In the single SNP analysis, 4 SNPs on BTA 7, 15 and 16 were detected when multiple testing was corrected for the family-wise error rate by the Bonferroni method ( $p < 0.05$  resulted in a Bonferroni threshold of  $p < 1.2e10^{-6}$ ). 270 significant SNPs, distributed over all the 29 *Bos taurus* autosomes, were detected when using the false discovery rate (FDR) with a  $q < 0.05$  level. When applying penalized multiple regression, we found 143 SNPs (distributed over large regions of the genome) using the elastic net (mixing parameter  $\alpha = 0.05$ ) in combination with 1000-fold cross-validation.

**Conclusions.** The analyses showed that single SNP and penalized multiple regression yielded different results. We know from extensive analyses of simulated data that the elastic net method with  $\alpha = 0.05$  has an almost optimal trade-off between Type I and Type II error rates in GWA studies where SNPs have moderate to high levels of linkage disequilibrium between each other. Therefore, it seems reasonable to draw the conclusion that single SNP analyses produce unreliable results, and that longevity in Fleckvieh is a character that probably is influenced by many genes.

# QTL detection and Marker Assisted Selection for the Resistance to *Phytophthora palmivora* using a F2 *Theobroma Cacao* L. progeny

Rogério Mercês Ferreira Santos<sup>1-2</sup>, Elisa Susilene Lisboa dos Santos<sup>3</sup>, Edna Dora Martins Newman Luz<sup>1</sup>, Mathilde Allegré<sup>4</sup>, Xavier Argout<sup>4</sup>, Olivier Fouet<sup>4</sup>, Karina Gramacho<sup>1</sup>, Dayse Santana<sup>2</sup>, João Marcos Mello Ribeiro<sup>1</sup>, Edaci Primo Da Costa<sup>1</sup>, Claire Lanaud<sup>4</sup>, Didier Clément<sup>1-4\*</sup>.

<sup>1</sup>Comissão Executiva do Plano da Lavoura Cacaueira, Centro de Pesquisas do Cacau, Rodovia Ilhéus-Itabuna, Km 12, Caixa Postal 07, CEP 45600-970 Itabuna, BA, Brasil

<sup>2</sup>Universidade Estadual Santa Cruz, 45650-000 Ilhéus, Bahia, Brazil

<sup>3</sup>Universidade Estadual do Sudoeste da Bahia, Campus Universitário Juvino Oliveira - Itapetinga - BA

<sup>4</sup>Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Avenue d'Agropolis, 34398, Cedex 5, Montpellier, France.

With the financial contribution of Agropolis Fondation

\*Presenting author: Didier Clément, email: didier.clement@cirad.fr

**Background.** Obtain sustainable varieties resistant to disease is one of the main goals of genetic improvement of cocoa. In partnership between CEPLAC (Comissão Executiva do Plano da Lavoura Cacaueira) and CIRAD (Centre International de Recherche Agronomique pour le Développement), studies to detect and characterize QTLs for resistance to witches' broom (*Moniliophthora perniciosa*) and to pod rot (*Phytophthora palmivora*), are conducted. QTL detection for resistance to *Phytophthora palmivora*, were carried out from a segregating F2 progeny (172 ind) resulting from the cross between Scavina-6 (resistant Forastero clone) and ICS1 (susceptible Trinitario clone). Phenotypic symptoms were revealed seven days after inoculation by a *P. palmivora* strain (CEPLAC phytopathology laboratory collection). A drop of 0.2ml of a concentrated solution of zoospores was deposited on leaf discs. Phenotypic analyses on necrosis severity (note 0 :without penetration point until note 5 :total necrosis) were obtained from sixty discs per genotype. QTL analyses were carried out using a genetic map established with 202 markers (SNPs and SSRs). The remaining part of the F2 population (972 individuals excluding plants used for the QTL analysis), was genotyped with the SSR markers associated with QTLs detected. The 972 genotypes were classified according to the presence of favorable QTL-allele or unfavorable QTL-allele. Phenotypic analyses on necrosis severity were carried under the same conditions as for the QTL analysis and two sets of thirty seven selected genotypes with favorable QTLs and unfavorable QTLs, were inoculated.

**Results.** The remaining part of the F2 population (972 individuals) was genotyped with the SSR markers : mTcCIR91, mTcCIR175 et mTcCIR430, respectively associated with the three QTLs detected in the linkage groups 10, 3 and 2, explaining 25% of the total variation. Statistical analysis results presented here, show a significative difference between both sets of plants, showing a very efficiency selection for *Phytophthora* resistance only with these markers close to QTLs.

**Conclusion.** Scavina-6 is also an important source of resistance to witches' broom (*Moniliophthora perniciosa*) The same approach using this same F2 population Scavina-6 by ICS1, is underway to select resistant plants.

G. M. Tahi, G. M. Tahi, B. I. Kébé, A. Sangare, F. Mondeil, C. Cilas and A. B. Eskes (2006). Foliar resistance of cacao (*Theobroma cacao*) to *Phytophthora palmivora* as an indicator of pod resistance in the field: interaction of cacao genotype, leaf age and duration of incubation. *Plant Pathology* 55, 776–782.

C. Lanaud, O. Fouet, D. Clément, M. Boccara, A. M. Risterucci, S. Surujdeo-Maharaj, T. Legavre, X. Argout (2009). A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. *Mol Breeding* DOI 10.1007/s11032-009-9297-4.

# Genome-wide association genetics to identify loci involved in responses to Potyviruses in plants

V. Schurdi-Levraud<sup>1\*</sup>, P. Cosson<sup>1</sup>, Z. Nehr<sup>1</sup>, M. Caballero<sup>1</sup>, C. Caranta<sup>2</sup>, F. Roux<sup>3</sup>, F. Revers<sup>1</sup>.

<sup>1</sup>UMR 1332 Biologie du Fruit et Pathologie, INRA Université de Bordeaux, avenue E. Bourleaux, F-33883 Villenave d'Ornon cedex.

<sup>2</sup>UR GAFL 1052, INRA Avignon, Domaine Saint-Maurice - BP 94, F-84143 Monfavet cedex.

<sup>3</sup>Université des Sciences et Technologies de Lille 1 Laboratoire de Génétique et Evolution des Populations Végétales - FRE CNRS 3268, F-59655 Villeneuve d'Ascq Cedex.

\*Presenting author: Valérie Schurdi-Levraud, email: vschurdi@bordeaux.inra.fr

**The aim of the project is to identify loci involved in susceptibility/resistance responses of *Arabidopsis thaliana* to Potyviruses using genome-wide association (GWA) mapping.**

The project will allow, at once, designing a map of loci involved in **quantitative response to two potyviruses, Turnip mosaic virus (TuMV) and Watermelon mosaic virus (WMV)** in the species *Arabidopsis thaliana*. It will be an important first step towards **the identification of candidate genes or quantitative traits loci (QTL)** responsible for quantitative response and the **cloning** of the most important of them. The project will also contribute to **an initial setting in understanding plant/virus co-evolution** through the **comparison of loci involved in the responses** to the two challenged viruses.

In **first steps**, robust phenotyping protocols are developed on a **narrow range of genotypes** known to reveal a large variability of responses to TuMV.

We optimize :

- Standardization of **plant culture**, growth conditions and inoculation conditions, **control of environment**,
- Defining **type of sample** through the following of TuMV-GusGFP replication and movement,
- Defining scoring **classes for symptoms**,
- Defining **sampling periods** through analysis of virus accumulation by RT-QPCR,
- Optimizing **mean-throughput sampling and RT-QPCR**.

Based on the first data, we will measure on the whole plant set : (1) **symptoms severity** and distribution using a range of scoring classes, (2) **virus accumulation** in inoculated and systemic (further growing non inoculated) leaves by RT-QPCR, (3) **infection kinetics** by sampling inoculated leaves and systemic leaves 5, 7, 9 and 12 days after the inoculation and performing RT-QPCR, (4) **length of floral hamp**s 21 days after inoculation, (5) **weight of total plant** 28 days after inoculation.

**Association between the phenotype and the genotype** will then be performed using a 250k SNPs database.

# The changes of marker-allele frequency induced by selection

Elena Orsini<sup>2\*</sup>, Chris C. Schön<sup>1</sup>, and Albrecht E. Melchinger<sup>2</sup>

<sup>1</sup>Plant Breeding, Center for Life and Food Sciences Weihenstephan, Technische Universität München, Germany

<sup>2</sup>Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany

\*Presenting author: Elena Orsini, email: e.orsini@uni-hohenheim.de

**Background.** The expression “selective genotyping” in QTL mapping studies, refers to the methods of detecting linkages between marker loci and QTL affecting the trait of interest when only part of the population is genotyped. This approach can be unidirectional if one of the phenotypic extremes of the trait distribution is analyzed or bidirectional, if both tails are considered. One of the consequences of selection is the change of the frequency of marker-alleles linked to QTL involved in the variation of the selected trait. Lines selected from the higher tail of the phenotypic distribution are expected to carry more favorable alleles compared to those with low phenotypic values. As a consequence, one method to detect QTL-marker linkage in selective genotyping is to examine marker-allele frequencies in selected lines. Such an approach is referred to as “trait based approach”. The frequencies of the marker-alleles can be compared between the two-tails in case of bidirectional selective genotyping or with the expected Mendelian segregation ratio in the case of unidirectional selective genotyping.

Although numerous analyses have been performed dealing with selective genotyping and trait based methods, nearly all of them focus on one QTL closely linked to a marker. Studies of multiple linked QTL concentrate mainly on bidirectional selective genotyping.

**Results and Conclusions.** In our work, we consider unidirectional selective genotyping and we analyze changes of marker-allele frequencies under arbitrary linkage and arbitrary epistasis between two QTL affecting the same trait or correlated traits. We propose a new trait based method for selective genotyping of a DH or a RIL population. Our mathematical expressions are derived from the probability density functions of the marker genotypes in the selected lines. The expected changes in marker-allele frequencies are shown to be a function of the linkage coefficient of Schnell (1961) and the truncation selection point.

*Schnell, F., 1961 Some general formulations of linkage effects in inbreeding. Genetics 46: 947-957.*

# Practical information

**Radio Taxi Rennais :**

Tel : 02 99 30 79 79 (24h/24h)

**Agrocampus Ouest :**

65 rue de Saint-Brieuc – CS84215 – F-35042 Rennes Cedex

**Taverne de la marine:**

2 place de bretagne, 35000 Rennes, France

Tel : 02 99 31 53 84

# Visit of the Brittany Parliament

Constitution of groups :

<b>Orange</b>	<b>Green</b>	<b>Yellow</b>	<b>Brown</b>
Olivier Demeure	Jean Michel Elsen	Olivier Filangi	Pascale Le Roy
Thomas Obadia	Xiaoqiang Wang	Anne Yuna Blum	Christine Gourbe
Rodolfo Juan Carlos Cantet	Koen Huvenaars	Christèle Robert Granié	Sandrine Schwob
Patrik Waldmann	Ghyslaine Schopen	Sophie Allais	Marie Noelle Fouilloux
Didier Clement	Abe Huisman	Nicolas Deschamp	François Guillaume
Chonglong Wang	Anouk Schurink	Félicien Shumbusho	Thierry Tribout
Zheya Sheng	Naomi Duijvesteijn	Anne Ricard	Marie Pierre Sanchez
Mohammad Mahdi Shariati	Marcos Soares Lopes	Gilles Charmet	Cédric Dufour
Naveen Kadri	Yanyan Sun	Hélène Muranty	Frédéric Hospital
Donghui MA	Suzanne Rowe	Carine Colombani	Florence Phocas
Joseph O Ogutu	Valentina Riggio	Valérie Schurdi-levraud	Vincent Ducrocq
Torben Schulz-Streeck	Oswald Matika	Patrick Cosson	Hélène Gilbert
Elena Orsini	Zivan Karaman	Simon Teyssedre	Sébastien Arico
Usai Graziano	Chris Hoze	Sébastien Ducrocq	Etienne Neuts
Magdalena Malina	Weronica Ek	Renaud Rincet	Romain Dassonneville
Anna Mucha	Ronald Nelson	Julie Hamon	Jean Pierre Bidanel
Tomasz Strabel	Mats Pettersson	Pascal Croiseau	Loys Bodin
Marcin Pszczola	Xia Shen	Eric Venot	Stéphane Nicolas
Sebastian Mucha	Marcin Kierczak	Trang Tran	Nicolas Bardol
Isaac Kofi Bimpong	Carl Nettelblad	Mohamed Kileh Wais	Guillaume Chapuis
Hélène Rome	Eliette Combes	Milagros Garcia	Pierre Cherel
Yvan Lebras	Huimin Yi		François Besnier
Carole Moreno	Liang Li		Damien Valour
Ingrid David	Dorian Garrick		Tom Druet
Christopher Sauvage	Julien Jacquin		

# List of Registered Participants

Title	last name	first name	Email	Country	Company
Miss	Allais	Sophie	sophie.allais@unceia.fr	France	UNCEIA
Mr	Bardol	Nicolas	nbardol@moulon.inra.fr	France	INRA
Mr	Besnier	Francois	francois.besnier@imr.no	Norway	IMR
Mr	Bidanel	Jean pierre	Jean-Pierre.Bidanel@jouy.inra.fr	France	INRA
Mr	Bimpong	Isaac kofi	kofibimpong@yahoo.com	Senegal	Africa Rice Sahel Regional Station
Mr	Bodin	Loys	Loys.Bodin@toulouse.inra.fr	France	SAGA
Mr	Cantet	Rodolfo juan carlos	rcantet@agro.uba.ar	Argentina	CONICET
Mr	Chapuis	Guillaume	guillaume.chapuis@irisa.fr	France	Irisa
Mr	Charmet	Gilles	gilles.charmet@clermont.inra.fr	France	INRA
Mr	Cherel	Pierre	pierre.cherel@hendrix-genetics.com	France	Hendrix Genetics
Mr	Clement	Didier	didier.clement@cirad.fr	Brazil	CIRAD
Miss	Colombani	Carine	carine.colombani@toulouse.inra.fr	France	INRA
Miss	Combes	Eliette	eliette.combes@limagrain.com	France	Limagrain Europe
Mr	Cosson	Patrick	pcosson@bordeaux.inra.fr	France	INRA
Mr	Croiseau	Pascal	pascal.croiseau@jouy.inra.fr	France	G2B
Mr	Dassonneville	Romain	romain.dassonneville@jouy.inra.fr	France	INRA
Mrs	David	Ingrid	Ingrid.David@toulouse.inra.fr	France	INRA
Mr	Dechamp	Nicolas	nicolas.dechamp@jouy.inra.fr	France	INRA
Mr	Ducrocq	Sebastien	sebastien.ducrocq@limagrain.com	France	Limagrain Europe
Mr	Dufour	Cedric	cedric.dufour@limagrain.com	France	LIMAGRAIN
Miss	Duijvesteijn	Naomi	diane.bolte@ipg.nl	The Netherlands	Institute for Pig Genetics B.V.
Miss	Ek	Weronica	weronica.ek@hgen.slu.se	Sweden	Uppsala University
Miss	Fouilloux	Marie-noëlle	marie-noelle.fouilloux@inst-elevage.asso.fr	France	Institut de L'Elevage
Mrs	Garcia	Milagros	milagros.garcia@toulouse.inra.fr	France	INRA
Miss	Gilbert	Hélène	helene.gilbert@toulouse.inra.fr	France	INRA
Mr	Guillaume	François	francois.guillaume@jouy.inra.fr	France	Institut de l'eleavage
Miss	Hamon	Julie	Julie.Hamon@polytech-lille.net	France	genes diffusion / INRIA
Mr	Hospital	Frederic	frederic.hospital@jouy.inra.fr	France	INRA
Miss	Hoze	Chris	chris.hoze@jouy.inra.fr	France	UNCEIA
Mr	Huisman	Abe	abe.huisman@hendrix-genetics.com	The Netherlands	Hendrix Genetics
Mr	Huvenaars	Koen	sa@keygene.com	The Netherlands	Agro Business Park
Mr	Jacquin	Julien	julien.jacquin@toulouse.inra.fr	France	INRA
Mr	Kadri	Naveen	Naveenk.kadri@agrsci.dk	Denmark	Aarhus University
Mr	Karaman	Zivan	zivan.karaman@limagrain.com	France	LIMAGRAIN
Mr	Kierczak	Marcin	marcin.kierczak@slu.se	Sweden	Uppsala University
Mr	Kileh wais	Mohamed	mohamedkileh@gmail.com	France	UMR SAGA / INRA
Mr	Le bras	Yvan	yvan.lebras@rennes.inra.fr	France	BioGenOuest
Mr	Li	Liang	liang.li@limagrainchina.com	France	LIMAGRAIN EUROPE
Mr	Ma	Donghui	dma@kws.com	Germany	KWS
Miss	Malina	Magdalena	malina@math.uni.wroc.pl	Poland	Wroclaw University
Mr	Matika	Oswald	oswald.matika@roslin.ed.ac.uk	United kingdom	Roslin Institute
Mrs	Moreno	Carole r	moreno@toulouse.inra.fr	France	UMR SAGA / INRA
Mrs	Mucha	Anna	anna.mucha@up.wroc.pl	Poland	Wroclaw University
Mr	Mucha	Sebastian	smucha@jay.up.poznan.pl	Poland	Poznan University
Mrs	Muranty	Helene	Helene.Muranty@orleans.inra.fr	France	UR AGPF / INRA
Mr	Nelson	Ronald	ronnie.nelson@slu.se	Sweden	Uppsala University
Mr	Nettelblad	Carl	carl.nettelblad@it.uu.se	Sweden	Uppsala University



## 15<sup>th</sup> QTL-MAS Workshop, May 19-20, 2011, Rennes, France

Mr	Neuts	Etienne	etienne.neuts@limousine.org	France	France Limousin Sélection
Mr	Nicolas	Stéphane	snicolas@moulon.inra.fr	France	UMR Génétique Végétale
Mr	Ogotu	Joseph o	jogutu2007@gmail.com	Germany	University of Hohenheim
Miss	Orsini	Elena	e.orsini@uni-hohenheim.de	Germany	University of Hohenheim
Mr	Pettersson	Mats	mats.pettersson@slu.se	Sweden	Uppsala University
Mrs	Phocas	Florence	florence.phocas@jouy.inra.fr	France	INRA
Mr	Pszczola	Marcin	mbee@jay.up.poznan.pl	Poland	Poznan University
Mrs	Ricard	Anne	anne.ricard@toulouse.inra.fr	France	INRA
Miss	Riggio	Valentina	valentina.riggio@roslin.ed.ac.uk	United kingdom	Roslin Institute
Mr	Rincent	Renaud	renaud.rincent@moulon.inra.fr	France	LIMAGRAIN EUROPE
Mrs	Robert-granie	Christèle	Christele.Robert-Granie@toulouse.inra.fr	France	UMR SAGA / INRA
Mrs	Rowe	Suzanne	suzanne.rowe@roslin.ed.ac.uk	United kingdom	Roslin Institute
Miss	Sanchez	Marie-pierre	marie-pierre.sanchez@jouy.inra.fr	France	INRA
Mr	Sauvage	Christopher	christopher.sauvage@avignon.inra.fr	France	INRA
Miss	Schopen	Ghyslaine	ghyslaine.schopen@crv4all.com	The netherlands	CRV
Mr	Schulz-streeck	Torben	torben.schulz-streeck@uni-hohenheim.de	Germany	University of Hohenheim
Mrs	Schurdi-levraud	Valérie	vschurdi@bordeaux.inra.fr	France	INRA
Miss	Schurink	Anouk	anoukschurink@hotmail.com	The netherlands	Wageningen University
Mrs	Schwob	Sandrine	sandrine.schwob@jouy.inra.fr	France	INRA
Mr	Shariati	Mohammad mahdi	mohammad.shariati@agrsci.dk	Denmark	Aarhus University
Mr	Shen	Xia	xia.shen@lcb.uu.se	Sweden	Uppsala University
Miss	Sheng	Zheyu	Zheyu.Sheng@slu.se	China	China Agricultural University
Mr	Shumbusho	Félicien	felicien.toulouse@toulouse.inra.fr	France	INRA
Mr	Soares lopes	Marcos	diane.bolte@ipg.nl	The netherlands	Institute for Pig Genetics B.V.
Mr	Strabel	Tomasz	strabel@jay.up.poznan.pl	Poland	Poznan University of Life Sciences
Miss	Sun	Yanyan	yanyan.sun@wur.nl	The netherlands	Wageningen University
Mr	Teyssedre	Simon	simon.teyssedre@toulouse.inra.fr	France	INRA
Mr	Tran	Trang	trang.tran@ingenomix.fr	France	INGENOMIX
Mr	Tribout	Thierry	thierry.tribout@jouy.inra.fr	France	INRA
Mr	Usai	Graziano	Graziano.usai@gmail.com	Italy	AGRI Sardegna
Mr	Valour	Damien	damien.valour@jouy.inra.fr	France	INRA BDR
Mr	Venot	Eric	eric.venot@jouy.inra.fr	France	INRA
Mr	Waldmann	Patrik	Patrik.Waldmann@boku.ac.at	Austria	BOKU
Mr	Wang	Chonglong	ahwchl@163.com	China	China Agricultural University
Miss	Yi	Huimin	huimin.yi@limagrain.com	France	LIMAGRAIN EUROPE