DETECTING QTL AND ESTIMATING BREEDING VALUES IN THE QTLMAS 2010 DATA

Albart Coster (albart.coster@wur.nl);
Mario Calus (mario.calus@wur.nl).

Animal Breeding and Genomics Centre

17-18 May 2010

EXPLORING THE DATA

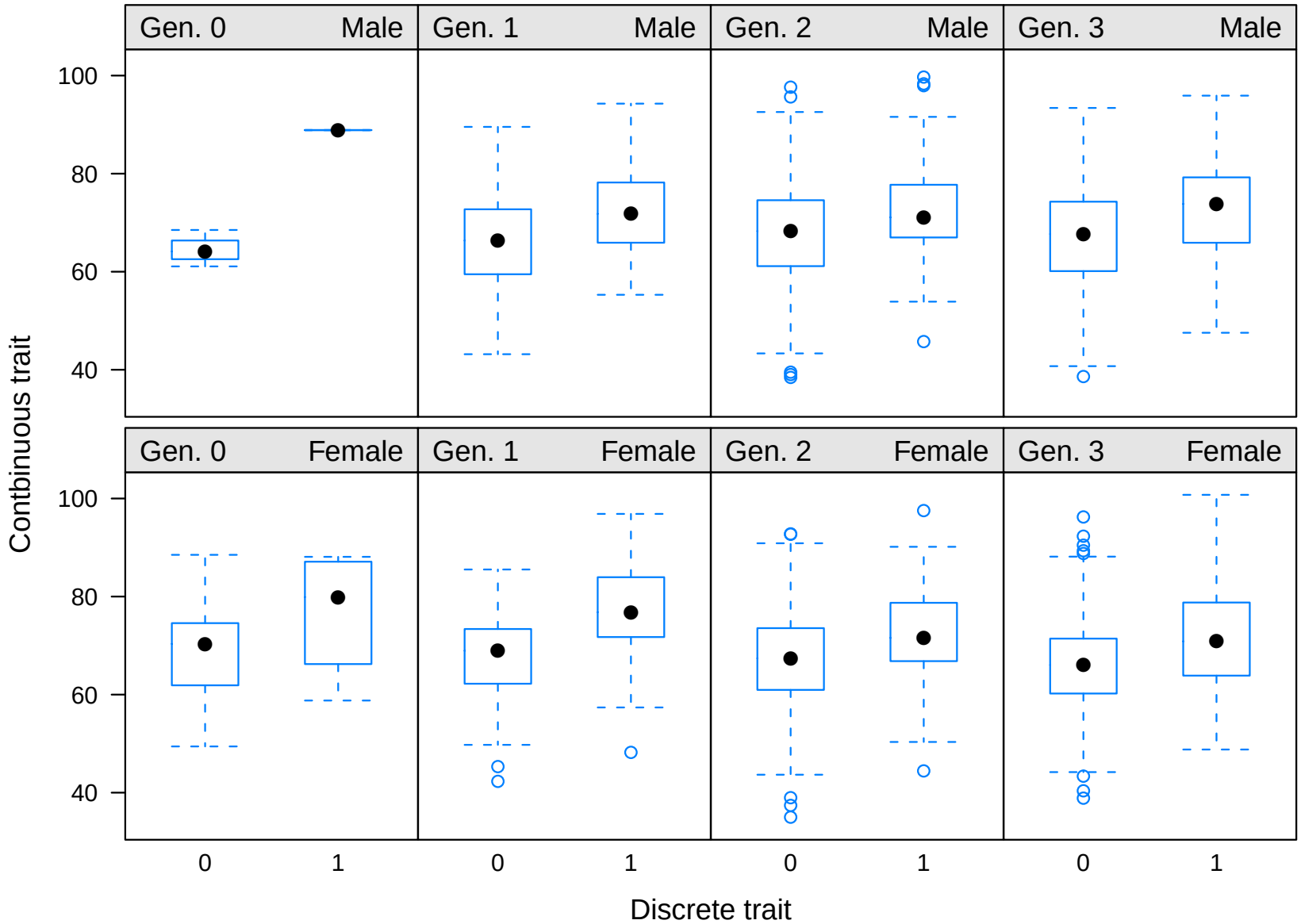
Data:

- Five generations of pedigree (gen 0 to 4);
- genotypes of approx. 10.000 markers for all individuals;
- phenotypes of a continuous and a discrete trait for individuals in generations 0 to 3.

Initial questions:

- Are there effects of sex or generation in the data?
- Are the two traits heritable?
- Are the two traits correlated?
- Is there a genetic correlation between the traits?

BOXPLOTS OF THE DATA



BIVARIATE ANIMAL MODEL

$$\mathbf{Y} = \mathbf{XB} + \mathbf{ZU} + \mathbf{E}$$

$$\mathbf{U} \sim \mathbf{N}\left(0, \mathbf{A} \otimes \begin{bmatrix} \sigma^2_{gc} & \sigma_{gc,gd} \\ \sigma_{gc,gd} & \sigma^2_{gd} \end{bmatrix}\right)$$

$$\mathbf{E} \sim \mathbf{N}\left(0, \mathbf{I} \otimes \begin{bmatrix} \sigma^2_{ec} & 0 \\ 0 & \sigma^2_{ed} \end{bmatrix}\right)$$

TABLE: Results of the multivariate animal model..

h^2 cont.	h^2 disc.	ρ_g	ρ_p
0.53 (0.06)	0.22 (0.04)	0.66 (0.09)	0.25 (0.03)

Conclusion initial analysis

- No sex effects was present.
- No selection was present.
- Traits are phenotypically correlated.
- Both traits are heritable; continuous trait more than discrete trait.
- Traits are genetically correlated.

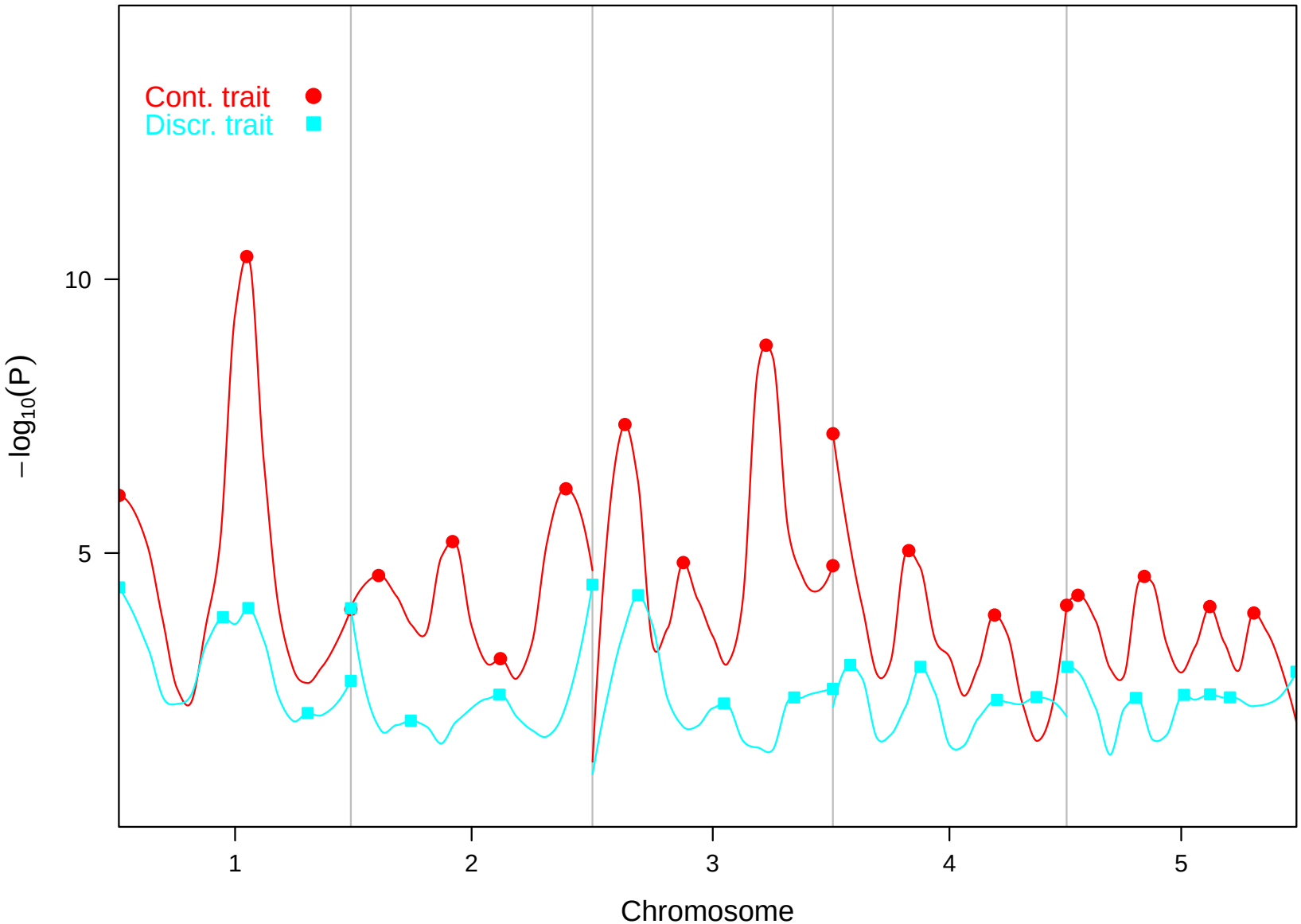
Objectives:

- Easy analysis of the marker data.
- Observe genomic regions of interest.
- Compare to more sophisticated method.

Model:

$$\mathbf{y} = \mathbf{G}\mathbf{v} + \mathbf{e}$$

SINGLE MARKER REGRESSION: RESULTS



X is matrix of independent variables (genotypes)
Y is matrix of dependent variables (phenotypes)

PLSR decomposes **X** and **Y** :

$$\mathbf{X} = \mathbf{T}\mathbf{W}'$$
$$\mathbf{Y} = \mathbf{U}\mathbf{Q}'$$

$$\mathbf{U} = \mathbf{B}\mathbf{T}$$

Requirements:

- both **T** and **U** orthogonal;
- maximal covariance between **T** and **U** ;
- **T** and **U** sequentially contain maximal variance of **X** and **Y** .

From: P. Geladi, *Anal Chim Acta* 185:1-17; De Jong *Chemom Intell Lab Syst* 18:251-263.

Application of PLSR to data: two step approach

Step 1 (finding QTL)

- Independent PLSR models for each chromosome.
- Bootstrapping to calculate empirical s.e. of regression coefficients for markers.
- Standardize regression coefficients of each marker, $\frac{\hat{\beta}}{\hat{se}}$.
- Detect QTL based on smoothed profile of standardized regression coefficients: QTL is a local maximum of the smoothed curve (first derivative = 0, second derivative < 0).

Step 2 (calculation of EBV)

- Use 1000 most significant marker for each trait: 2000 markers in final model.
- Fit PLSR model with 2000 markers.
- Use regression coefficients to calculate EBV of individuals for both traits.

SMOOTHED CURVE OF STANDARDIZED β'S

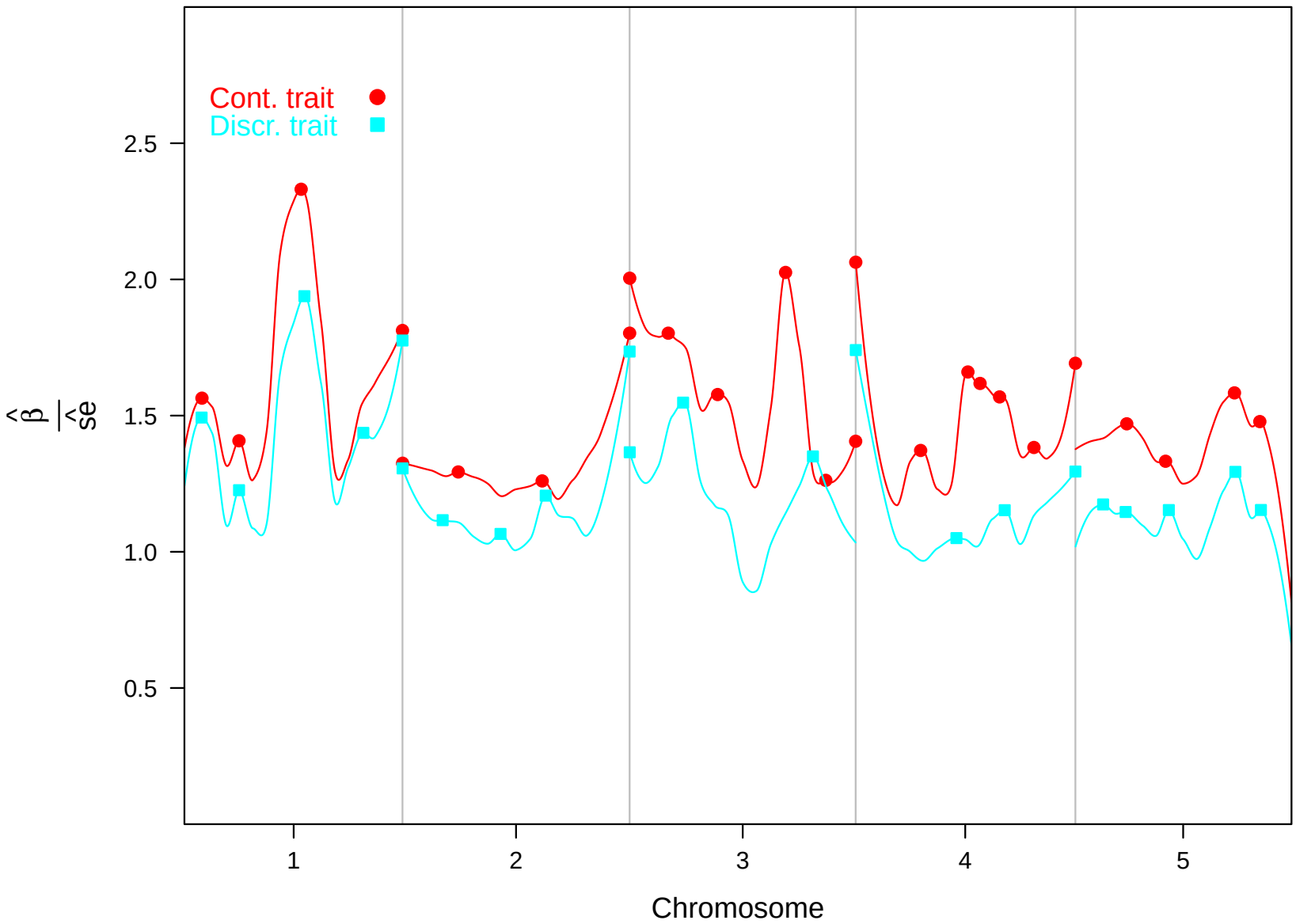


TABLE: Table of QTL detected on chrom. 1 and 3.

marker		Cont. trait		Discr. trait	
	MAF	$\hat{\beta}$	$\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_g}$	$\hat{\beta}$	$\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_g}$
Chrom. 1					
156	0.46	2.9920	0.0817	**0.0612**	**0.0405**
159	0.39	**-2.3459**	**0.0479**	-0.0654	0.0442
494	0.48	**0.3242**	**0.0010**	-0.0369	0.0148
495	0.06	-1.4611	0.0046	**0.0672**	**0.0116**
1058	0.31	**4.1433**	**0.1355**	0.0426	0.0170
1087	0.45	-0.4640	0.0020	**-0.0397**	**0.0170**
1621	0.41	0.2338	0.0005	**0.0108**	**0.0012**
1976	0.29	**-0.0818**	**0.0001**	**0.0336**	**0.0102**
Chrom. 3					
4035	0.31	**-1.6002**	**0.0200**	**-0.0150**	**0.0021**
4384	0.37	**-1.3395**	**0.0153**	-0.0650	0.0427
4519	0.40	-0.1875	0.0003	**-0.0489**	**0.0249**
4832	0.47	**-1.7533**	**0.0281**	-0.0576	0.0360
5447	0.45	**1.9137**	**0.0333**	0.0062	0.0004
5695	0.46	-1.2392	0.0140	**0.0278**	**0.0084**
5811	0.38	**-2.1743**	**0.0407**	-0.0930	0.0883
6082	0.28	**0.6369**	**0.0030**	0.0123	0.0013

TABLE: Correlation between phenotype and EBV summarized per generation.

Trait	0	1	2	3
Cont. trait	0.66	0.70	0.69	0.69
Discr. trait	0.64	0.62	0.55	0.54

Conclusions

- PLSR is a good alternative method for finding QTL and calculating EBV.
- Some marker selection is required for large numbers of markers.
- Smoothing curves of P-values of regression coefficients is an intuitive graphical method to find QTL.

- Several QTL were found for both trait, some of them pleiotropic.
- Expect that many pleiotropic QTL were simulated.