

RESEARCH

Open Access

# Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs

Tage Ostensen<sup>1\*</sup>, Ole F Christensen<sup>2\*</sup>, Mark Henryon<sup>1,3</sup>, Bjarne Nielsen<sup>1</sup>, Guosheng Su<sup>2</sup> and Per Madsen<sup>2</sup>

## Abstract

**Background:** Genomic selection can be implemented by a multi-step procedure, which requires a response variable and a statistical method. For pure-bred pigs, it was hypothesised that deregressed estimated breeding values (EBV) with the parent average removed as the response variable generate higher reliabilities of genomic breeding values than EBV, and that the normal, thick-tailed and mixture-distribution models yield similar reliabilities.

**Methods:** Reliabilities of genomic breeding values were estimated with EBV and deregressed EBV as response variables and under the three statistical methods, genomic BLUP, Bayesian Lasso and MIXTURE. The methods were examined by splitting data into a reference data set of 1375 genotyped animals that were performance tested before October 2008, and 536 genotyped validation animals that were performance tested after October 2008. The traits examined were daily gain and feed conversion ratio.

**Results:** Using deregressed EBV as the response variable yielded 18 to 39% higher reliabilities of the genomic breeding values than using EBV as the response variable. For daily gain, the increase in reliability due to deregression was significant and approximately 35%, whereas for feed conversion ratio it ranged between 18 and 39% and was significant only when MIXTURE was used. Genomic BLUP, Bayesian Lasso and MIXTURE had similar reliabilities.

**Conclusions:** Deregressed EBV is the preferred response variable, whereas the choice of statistical method is less critical for pure-bred pigs. The increase of 18 to 39% in reliability is worthwhile, since the reliabilities of the genomic breeding values directly affect the returns from genomic selection.

## Background

Genomic selection in pure-bred pigs can be implemented using a multi-step procedure. Effects of dense genetic markers are estimated using a reference population and these effects are used to predict genomic breeding values (GBV) of selection candidates [1]. Implementing a multi-step procedure relies on two prerequisites: 1) a response variable that summarises the genetic information for reference animals, and 2) a statistical method that associates the response variable to the marker information. The choice of response variable

and statistical method may well depend on the data structure. Therefore, the challenge is to find a suitable response variable and statistical method that can handle pure-bred pig data.

Pure-bred pig data often have low and varying reliabilities of estimated breeding values (EBV). Some Duroc pigs in the Danish breeding scheme only have their own records, others only offspring records, while some have both - depending on the trait. Furthermore, only a subset of pigs are genotyped and these pigs tend to be closely related. Therefore, a response variable and a statistical method capable of handling such data are needed.

Among the possible response variables, EBV and deregressed EBV are the most promising to date. Garrick et al. [2] showed, at least in theory, that deregressed EBV with the parent average removed (hereafter referred

\* Correspondence: tao@lf.dk; OleF.Christensen@agrsci.dk

<sup>1</sup>The Danish Agricultural and Food Council, Pig Research Centre, Breeding and Genetics, Axeltorv 3, DK-1609, Denmark

<sup>2</sup>Aarhus University, Faculty of Agricultural Sciences, Department of Genetics and Biotechnology, Blichers Allé 20, P.O. BOX 50, DK-8830, Tjele, Denmark  
Full list of author information is available at the end of the article

to as “deregressed EBV”) yield more accurate genomic breeding values than EBV for two reasons. First, deregressed EBV as the response variable results in less double-counting compared to EBV, because the proposed deregressed EBV excludes ancestral information. If both an offspring and its parent are genotyped, the degree of double-counting decreases when using deregressed EBV as the response variable. Second, using EBV as the response variable results in double shrinkage of the genomic breeding values, particularly when the reliabilities of the EBV are low. In dairy cattle, Guo et al. [3] compared the two response variables on simulated data and found that deregressed EBV yielded slightly lower reliabilities. However, their simulated data were characterized by low degrees of double-counting. It appears that most pure-bred pig data, with EBV with low reliabilities and high degrees of double-counting, compared to Guo et al. [3], are more in line with the theoretical expectations. Therefore, this suggests that deregressed EBV as the response variable may yield higher reliabilities of genomic breeding values than EBV.

Three types of statistical models have been widely used in the literature [1,4-14]. These are models that assume the marker effects to be normally distributed, models that assume a thick-tailed distribution of marker effects, and models that assume a mixture of two distributions. The performance of the different models is predominantly affected by the number of QTL, the marker density and the genetic relatedness of the population. However, there is no unambiguous evidence that one model will yield more accurate genomic breeding values for pure-bred pig data for two reasons. First, mixture-distribution models have shown promising results for some traits, in particular traits under weak selection, possibly because these traits can be influenced to a greater extent by a few large QTL [9,15]. Second, low genetic relatedness between reference animals and validation animals appears to favour mixture-distribution models [14], presumably because mixture-distribution models utilize linkage disequilibrium more efficiently than normal and thick-tailed distribution models. Probably for the same reason, Hayes et al. [9] concluded that thick-tailed models and models assuming normality were equally good only when data from a single cattle breed were analysed. It appears that for pure-bred pig data, for which there is strong selection on traits of interest and high relatedness between genotyped animals, none of the models would be favoured over the others. Therefore, the models assuming mixture-distribution, thick-tailed distribution and normal distribution might yield similar reliabilities of genomic breeding values.

In summary, we reasoned that 1) deregressed EBV as the response variable yields higher reliabilities of the GBV compared to EBV, and that 2) normal, thick-tailed

and mixture-distribution models yield similar reliabilities. To test these hypotheses, the three models and the two response variables were assessed for the reliabilities by which they could predict GBV for the two traits, daily gain and feed conversion ratio, in Danish Duroc pigs.

## Methods

### Procedure

Two response variables, EBV and deregressed EBV, and three statistical methods with normal, thick-tailed, and mixture-distributions, were assessed for their reliability to predict GBV for daily gain and feed conversion ratio in Danish Duroc pigs. The reliabilities were computed by splitting the genotyped animals into 1375 reference animals that were performance tested before October 2008 and 536 validation animals that were performance tested after October 2008 (further details are given below).

### Data

The Duroc pigs were part of the genetic evaluation system in Denmark. All data were supplied by the Danish Agriculture and Food Council, Pig Research Centre.

### Genotyping

A total of 1911 Danish Duroc pigs were genotyped using the Illumina PorcineSNP60 BeadChip (Illumina, San Diego, CA). A total of 26142 SNP markers and each of the animals met the following requirements. Each animal had a call rate greater than 0.95. Each marker was mapped to an autosome, had a minor-allele frequency greater than 0.05, a call-frequency score greater than 0.95, and a heterozygote frequency that did not deviate from Hardy-Weinberg expectations by more than  $1/4\sqrt{pq}$ , where  $p$  and  $q$  are the allele frequencies at the marker. The  $1/4\sqrt{pq}$  corresponds to a  $1/4$  standard-deviation unit when assuming a binomial distribution of alleles. For each animal and SNP-genotype combination, the GenCall score was greater than 0.65. Genotypes less than 0.65 were defined as missing. Animals with missing genotypes were allocated the population mean for the missing markers.

### Performance test and pedigree

Daily gain (g/day) and feed conversion ratio (feed units/kg gain) were observed in the interval 30-100 kg live weight. Recordings were performed in the period 1992 to 2010. The pedigrees of the animals were traced back to 1984, and consisted of 345686 and 52537 animals for daily gain and feed conversion ratio, respectively. Both pedigrees included 373 unknown parents (base animals).

The reference data consisted of available records on October 1st 2008 and included 313068 and 23628 measurements of daily gain and feed conversion ratio, respectively. All of the 1375 genotyped animals in the reference data had their own records for daily gain,

whereas only 898 genotyped animals had their own records for feed conversion ratio. There were 680 genotyped animals in the reference data that had more than four offspring records for daily gain and 633 genotyped animals that had more than four offspring records for feed conversion ratio.

The full data consisted of records available until May 2010 and included the records from the reference data and additional 32618 and 3344 measurements of daily gain and feed conversion ratio, respectively. All 536 genotyped validation animals were phenotyped for both daily gain and feed conversion ratio after October 1st, 2008.

### Response variables

The EBV were calculated for daily gain and feed conversion ratio by single-trait animal models, and deregressed EBV were calculated by applying the procedure proposed by Garrick et al. [2].

The single-trait animal models were based on the routine evaluation model, with the regression effect of start weight, fixed effect of herd-week-section, random effect of pen and a random additive genetic effect. The model fitted to daily gain also included a fixed effect for sex and a random effect for birth litter. The variance components were estimated using REML. The heritabilities for daily gain and feed conversion ratio were 0.27 and 0.21. The 1375 reference animals had mean reliabilities for EBV of 0.62 ( $sd = 0.18$ ) and 0.36 ( $sd = 0.12$ ) for daily gain and feed conversion ratio, respectively. The software DMU [16] was used to estimate variances and predict the breeding values. Single-trait animal models were used because preliminary analyses showed that predictions of GBV were not improved with a bivariate model.

The deregression procedure of Garrick et al. [2] adjusts for ancestral information, such that the deregressed EBV only contains their own and the descendant's information on each animal. The deregression also eliminates shrinkage contained in the EBV, and therefore deregressed EBV behave as though they were observations with a heritability equal to the reliability of the deregressed EBV (reliabilities computed as in Garrick et al. [2]). Deregressed EBV have unequal variances and should be used in a weighted analysis. To ensure the quality of deregressed EBV, only animals with a deregressed reliability above 0.05 were included in the analysis. This resulted in 35 animals being removed from the analyses with deregressed EBV for feed conversion ratio.

### Statistical methods to estimate marker effects

The marker effects were estimated by fitting linear, additive models to the response variables. The model was:

$$Y = \mu + X\beta + \varepsilon, \quad (1)$$

where  $Y$  is a  $n \times 1$  vector of responses, with  $n$  being the number of reference animals. The mean is denoted  $\mu$ , which is a scalar. The coefficient matrix for the markers,  $X$ , assume the values -1, 0 or 1, and has dimension  $n \times p$ , where  $p$  is the number of markers. The vector of marker effects is denoted  $\beta$  and has dimension  $p \times 1$ , whereas the vector of residuals,  $\varepsilon$ , has dimension  $n \times 1$ , and is distributed  $\varepsilon \sim N(0, \sigma_\varepsilon^2 W)$ , where  $W$  is a diagonal matrix with elements  $w_1, \dots, w_n$ .

For the analyses with EBV as the response variable,  $w_1 = \dots = w_n = 1$ , whereas for the analyses with deregressed EBV as the response variable, a weighted analysis was performed according to Garrick et al. [2]. The weight for the  $i$ th animal was  $w_i = (1 - h^2) / [(c + (1 - r_i^2) / r_i^2) h^2]$ , where  $c$ , the part of the genetic variance not explained by markers, was assumed to be 0.1,  $h^2$  was the heritability of the trait, and  $r_i^2$  was the reliability of the deregressed EBV of the  $i$ th animal.

The statistical models studied differ by the distributional assumptions made on  $\beta$ , and by the shortness of presentation they are referred to by the corresponding inferential procedure. A brief description of each method is provided in the following.

### GBLUP

The model for genomic BLUP (GBLUP) can be described as in equation (1), where the vector of marker effects are distributed  $\beta \sim N(0, \sigma_v^2 I)$ . Alternatively, the model could be written as  $Y = \mu + g + \varepsilon$ , where the genetic effects  $g \sim N(0, \sigma_v^2 XX^T)$ , i.e. a model with a genomic relationship matrix [7]. In this study, the latter form was used, where  $XX^T$  for computational reasons was replaced by  $(X - P)(X - P)^T$  with  $P$  containing the row means of  $X$  across animals. Variance components were estimated using REML, and the GBV were BLUP solutions from the mixed model equations. Software DMU [16] was used to fit GBLUP models.

### Bayesian Lasso

The model for Bayesian Lasso assumes that the marker effects follow a double exponential distribution [12,17], which is a thick-tailed distribution. This is also referred to as a Laplace distribution. Thus,  $\beta \sim \text{Laplace}(0, \lambda I)$  where  $\lambda$  is a scaling parameter. The MCMC software BayZ [18] was used for estimation. A Gibbs sampler was applied and a total of 40000 iterations of sampling was performed, with a burn-in of 10000 iterations.

### MIXTURE

The model for MIXTURE assumes that the marker effects follow a mixture of two normal distributions with similar characteristics as the mixture model in Meuwissen [19]. The distributional assumption can be described as:  $\beta \sim \pi_0 N(0, \sigma_{\beta_0}^2 I) + \pi_1 N(0, \sigma_{\beta_1}^2 I)$ , where  $\pi_0 = 0.9$ ,  $\pi_1 = 0.1$ , and  $\sigma_{\beta_0}^2 = 0.01 \cdot \sigma_{\beta_1}^2$ . The MCMC

software BayZ [18] was used for estimation. A Metropolis-Hastings sampler was applied with a total of 40000 iterations of sampling and a burn-in of 10000 iterations.

### Evaluation criterion

The predictive ability of each method was evaluated by the reliability of the GBV on the genotyped validation animals. To avoid use of overlapping information between the reference and validation animals [20], we based the validation on the own phenotype of genotyped validation animals, adjusted for fixed effects and non-genetic random effects. The computation of adjusted phenotypes for genotyped validation animals was based on the full data, and the adjusted phenotypes were the sum of EBV and the estimated residual errors ( $\hat{g} + \hat{e}$ ). The reliability of the GBV,  $r^2_{(GBV)}$ , is the squared correlation  $r^2_{(GBV,g)}$  between GBV and the vector of genetic effects,  $g$ , for genotyped validation animals. Using that the vector of residual errors,  $e$ , for these animals is independent of the GBV, and that  $g + e \approx \hat{g} + \hat{e}$ , we obtain a formula for the reliability of the GBV

$$r^2_{(GBV)} = r^2_{(GBV,g)} = r^2_{(GBV,g+e)} \cdot \omega \approx r^2_{(GBV,\hat{g}+\hat{e})} \cdot \omega, \quad (2)$$

where  $\hat{g}$  is the vector of EBV for validation animals,  $\hat{e}$  is the estimated residual vector,  $\omega = (\sigma_e^2 + \sigma_a^2)/\sigma_a^2$ ,  $\sigma_e^2$  is the residual variance and  $\sigma_a^2$  is the additive genetic variance of the trait. The formula above for the reliability is similar to other formulas shown in the literature [14,21].

A test was performed to investigate whether the reliabilities of the GBV from the two methods were significantly different from each other within each trait. The test hypothesis was that the reliability of GBV from method A was equal to the reliability of GBV from method B, i.e.  $H_0 : r^2_{(GBV_A,\hat{g}+\hat{e})} \cdot \omega = r^2_{(GBV_B,\hat{g}+\hat{e})} \cdot \omega$ . Since  $\omega$  is simply a constant for each trait, the test hypothesis can be reduced to  $H_0 : r_{(GBV_A,\hat{g}+\hat{e})} = r_{(GBV_B,\hat{g}+\hat{e})}$ . However,  $\hat{g} + \hat{e}$  appears in both correlations and the two correlations are not independent within each trait. This issue of testing equality of correlated correlations was addressed by the Hotelling-Williams  $t$ -test [22,23], which was applied to each trait with a confidence level of 5%.

## Results

### Response variables

The response variable, deregressed EBV, resulted in higher reliabilities than EBV for both feed conversion ratio and daily gain (Table 1). For daily gain, the reliabilities were on average 35% higher for deregressed EBV

with GBLUP, Bayesian Lasso and MIXTURE. The reliabilities for daily gain increased from approximately 0.25 to 0.34. For feed conversion ratio, the reliabilities were on average 18% higher for deregressed EBV for GBLUP and Bayesian Lasso, which approached significance (p-values between 0.05 and 0.11). Use of deregressed EBV instead of EBV increased reliabilities from approximately 0.16 to 0.19 for GBLUP and Bayesian Lasso. For the method MIXTURE, the deregressed EBV yielded 39% higher reliabilities than EBV, increasing reliabilities from 0.15 to 0.20.

### Statistical methods

The three statistical methods, GBLUP, Bayesian Lasso and MIXTURE did not yield different reliabilities (Table 1). For daily gain, the reliabilities ranged from 0.33 to 0.34 for deregressed EBV and from 0.25 to 0.26 for EBV. For feed conversion ratio, the reliabilities ranged from 0.19 to 0.20 for deregressed EBV and from 0.15 to 0.16 for EBV.

### Discussion

The hypothesis that deregressed EBV yield higher reliabilities than EBV was supported. The increase in reliability of 18 to 39% for Duroc pigs is worthwhile, since the reliabilities of the genomic breeding values directly affect the returns from genomic selection. The hypothesis that the different statistical methods yielded similar reliabilities was also supported. Therefore, we believe that when applying genomic selection within a purebred pig population, deregressed EBV is the preferred response variable, while the choice of the statistical method is less critical.

The fact that the increase in reliabilities due to deregression was so large was surprising, since results reported by Guo et al. [3] suggested that EBV is the more suitable response variable. There are three possible

**Table 1 Reliabilities of GBV,  $r^2_{(GBV)}$  based on the three statistical methods and the two response variables for daily gain and feed conversion ratio.**

	$r^2_{(GBV)}$	
	EBV	deregressed EBV
<b>Daily gain<sup>‡</sup></b>		
GBLUP	0.26 <sup>a</sup>	0.33 <sup>b</sup>
Bayesian Lasso	0.25 <sup>a</sup>	0.34 <sup>b</sup>
MIXTURE	0.25 <sup>a</sup>	0.34 <sup>b</sup>
<b>Feed conversion ratio<sup>‡</sup></b>		
GBLUP	0.16 <sup>ab</sup>	0.19 <sup>ab</sup>
Bayesian Lasso	0.16 <sup>ab</sup>	0.19 <sup>ab</sup>
MIXTURE	0.15 <sup>a</sup>	0.20 <sup>b</sup>

<sup>‡</sup> Reliabilities within daily gain and feed conversion ratio with different superscripts are significantly different ( $P < 0.05$ ) Reliabilities for daily gain and feed conversion ratio are not comparable

reasons. In our study less information was available for the reference animals compared to Guo et al. [3], which implies that the expected double shrinkage was increased when using EBV as the response variable. The amount of information was more heterogeneous among the reference animals, and this favours deregressed EBV. The heterogeneity in amount of information was larger for daily gain than for feed conversion ratio, which also explains why the effect of deregression was larger for this trait. The combination of less information per reference animal and several parent-offspring relationships within the reference population increases double-counting further. If a genotyped parent has only one offspring, and this offspring is genotyped, the degree of double-counting is much larger than if both the parent and the offspring have many ungenotyped offspring. So, the combination of sire-son relationships with a low and heterogeneous amount of information in the reference data seems to favour deregressed EBV over EBV.

There are two reasons why the three statistical methods performed equally well. First, the traits have been subject to strong selection, which suggests that MIXTURE would not have an advantage over the other methods [9,15]. Second, we considered only one breed, in which the genotyped animals were closely related, implying that Bayesian Lasso and MIXTURE, which utilize linkage disequilibrium more efficiently, have no advantage over GBLUP [9,14]. Thus, the small difference between the statistical methods is caused by the pure-bred pig data with its high relatedness between genotyped animals and traits that have been subject to strong selection.

The results from this study may only partly apply to other species, traits and data structures. The positive effect of using deregressed EBV as the response variable will depend on the degree of double-counting, and the amount and heterogeneity of information for genotyped animals. However, since the deregressed EBV proposed by Garrick et al. [2] are theoretically more appropriate than EBV, we believe that deregressed EBV will be advantageous in most circumstances. In contrast, the conclusion about equal performances of the statistical methods is not generic. For traits controlled by large QTL, data from multiple breeds or distantly related genotyped animals, or when using a denser marker panel, MIXTURE and Bayesian Lasso could outperform GBLUP. The reason for this is that MIXTURE and Bayesian Lasso utilize linkage disequilibrium more efficiently. We believe that the positive effect of using deregressed EBV applies to most situations, whereas choosing the best statistical method is more sensitive to the particular situation.

The applied multi-step method has the benefit of being readily applicable in most breeding schemes. EBV with reliabilities are easily predicted, and can then be

used to calculate deregressed EBV [2]. The deregression procedure itself is very simple and only requires little computing power. The association of marker effects with the response variable by GBLUP is relatively simple, since GBLUP is implemented in common breeding value estimation software. Bayesian Lasso and MIXTURE are available in various MCMC software, although particularly MIXTURE is sensitive to the prior information given, which makes the use of these methods less appealing. A final step, which has not been examined in the present study, would be to combine the genomic breeding values with the traditional pedigree information [24,25].

#### Acknowledgements

Two anonymous reviewers are thanked for their corrections. The authors gratefully acknowledge Luc Janss for help on the use of BayZ software, and David Lindsay for guidance on the structure of the paper. The work was funded by the Danish Ministry of Food, Agriculture and Fisheries (grant no. 3412-09-02505).

#### Author details

<sup>1</sup>The Danish Agricultural and Food Council, Pig Research Centre, Breeding and Genetics, Axeltorv 3, DK-1609, Denmark. <sup>2</sup>Aarhus University, Faculty of Agricultural Sciences, Department of Genetics and Biotechnology, Blichers Allé 20, P.O. BOX 50, DK-8830, Tjele, Denmark. <sup>3</sup>School of Animal Biology, University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia.

#### Authors' contributions

TO carried out most calculations and drafted the main parts of the manuscript, OFC created the EBVs from the animal model, made the GBLUP predictions, wrote significant parts of the manuscript and guided on statistical issues. MH guided in the structure of the manuscript, constructed the SNP data and wrote parts of the manuscript. BN, GS and PM helped conceive of the study and guided in the statistical analyses. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 1 March 2011 Accepted: 9 November 2011

Published: 9 November 2011

#### References

1. Meuwissen THE, Hayes BJ, Goddard ME: Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 2001, **157**:1819-1829.
2. Garrick DJ, Taylor JF, Fernando RL: Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 2009, **41**:55.
3. Guo G, Lund MS, Zhang Y, Su G: Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. *J Anim Breed Genet* 2010, **127**:423-432.
4. Xu S: Estimating polygenic effects using markers of the entire genome. *Genetics* 2003, **163**:789-801.
5. Roso VM, Schenkel FS, Miller SP, Schaeffer LR: Estimation of genetic effects in the presence of multicollinearity in multibreed beef cattle evaluation. *J Anim Sci* 2005, **83**:1788-1800.
6. Gianola D, Fernando RL, Stella A: Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 2006, **173**:1761-1776.
7. VanRaden PM: Efficient methods to compute genomic predictions. *J Dairy Sci* 2008, **91**:4414-4423.
8. Pimentel E, König S, Schenkel F, Simianer H: Comparison of statistical procedures for estimating polygenic effects using dense genome-wide marker data. *BMC Proceedings* 2009, **3**(Suppl 1):S12.

9. Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME: **Accuracy of genomic breeding values in multi-breed dairy cattle populations.** *Genet Sel Evol* 2009, **41**:51.
10. Su G, Guldbandsen B, Gregersen VR, Lund MS: **Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population.** *J Dairy Sci* 2010, **93**:1175-1183.
11. Cole JB, VanRaden PM, O'Connell JR, Van Tassel CP, Sonstegard TS, Schnabel RD, Taylor JF, Wiggans GR: **Distribution and location of genetic effects for dairy traits.** *J Dairy Sci* 2009, **92**:2931-2946.
12. de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM: **Predicting quantitative traits with regression models for dense molecular markers and pedigree.** *Genetics* 2009, **182**:375-385.
13. Gianola D, de los Campos G, Hill W, Manfredi E, Fernando R: **Additive genetic variability and the Bayesian alphabet.** *Genetics* 2009, **183**:347-363.
14. Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G: **The impact of genetic relationship information on genomic breeding values in German Holstein cattle.** *Genet Sel Evol* 2010, **42**:5.
15. Goddard ME, Hayes BJ, Meuwissen THE: **Genomic selection in farm animal species - lessons learnt and future perspectives.** In *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production. 1-6 August, 2010, Leipzig* Edited by: German Society for Animal Science 2010, paper 0701 E [http://www.kongressband.de/wcgalp2010/assets/pdf/0701.pdf].
16. Madsen P, Jensen J: **A users guide to DMU, version 6, release 4.7.** Manual, Faculty of agricultural science, University of Aarhus; 2008.
17. Yi N, Xu S: **Bayesian LASSO for quantitative trait loci mapping.** *Genetics* 2008, **179**:1045.
18. Janss LLG: *BayZ manual version 2.02* 2010 [http://bayz.biz/].
19. Meuwissen T: **Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping.** *Genet Sel Evol* 2009, **41**:35.
20. Amer PR, Banos G: **Implications of avoiding overlap between training and testing data sets when evaluating genomic predictions of genetic merit.** *J Dairy Sci* 2010, **93**:3320-3330.
21. Legarra A, Robert-Granié C, Manfredi E, Elsen JM: **Performance of genomic selection in mice.** *Genetics* 2008, **180**:611-618.
22. Dunn OJ, Clark V: **Comparison of tests of the equality of dependent correlation coefficients.** *J Amer Statistical Assoc* 1971, **66**:904-908.
23. Revelle W: *psych. Procedures for Psychological, Psychometric, and Personality Research* Northwestern University, Evanston, Illinois; 2010 [http://personality-project.org/r/psych.manual.pdf], [R package version 1.0-90].
24. Ducrocq V, Liu Z: **Combining genomic and classical information in national BLUP evaluations.** *Interbull Bull* 2009, **40**:172-177.
25. VanRaden PM, Van Tassel CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS: **Invited review: reliability of genomic predictions for North American Holstein bulls.** *J Dairy Sci* 2009, **92**:16-24.

doi:10.1186/1297-9686-43-38

**Cite this article as:** Ostensen et al.: Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genetics Selection Evolution* 2011 **43**:38.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

