



HAL
open science

Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses

Anne Ricard, Sophie S. Danvy, Andres Legarra

► To cite this version:

Anne Ricard, Sophie S. Danvy, Andres Legarra. Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses. *Journal of Animal Science*, 2013, 91 (3), pp.1076-1085. 10.2527/jas.2012-5256 . hal-01000368

HAL Id: hal-01000368

<https://hal.science/hal-01000368>

Submitted on 29 May 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.

JOURNAL OF ANIMAL SCIENCE

The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses

A. Ricard, S. Danvy and A. Legarra

J ANIM SCI 2013, 91:1076-1085.

doi: 10.2527/jas.2012-5256 originally published online December 10, 2012

The online version of this article, along with updated information and services, is located on
the World Wide Web at:

<http://www.journalofanimalscience.org/content/91/3/1076>



American Society of Animal Science

www.asas.org

Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses¹

A. Ricard,^{*2} S. Danvy,[†] and A. Legarra[‡]

^{*}INRA, UMR 1313, 78352 Jouy-en-Josas, France; [†]IFCE, Recherche et Innovation, 61310 Exmes, France; and [‡]INRA, UR 631, 31326 Castanet-Tolosan, France

ABSTRACT: Genomic evaluations often use as pseudo-phenotypes corrected means of progeny performances, like daughter yield deviations (DYD) in dairy species. In horse breeding, own performances are also available and performances from other relatives (as half sibs) may play an important part in the EBV because the number of progeny remains low, even for stallions. The first step for genomic selection in horses is therefore to generate pseudo-phenotypes for genomic analysis when parental or own information is considered. This work presents an easy method to compute deregressed EBV from regular pedigree-based genetic evaluations (EBV, reliabilities) to be used in genomic evaluations. The proposed methodology builds deregressed proofs so that they combine own performances (from genotyped individuals) and performances of relatives (outside of the genotyped sample). An application to show jumping horse data is presented. A sample of 908 stallions specialized in show jumping [71% Selle Français (SF), 17% foreign sport horses (FH), 13% Anglo Arab (AA)]

were genotyped. Genotyping was performed using the Illumina Equine SNP50 BeadChip, and after quality tests, 44,444 SNP were retained. Two methods were used for genomic evaluation: GBLUP and BayesC π , and 6 validation data sets were compared, chosen according to breeds SF + FH + AA or SF + FH, family structure (more than 3 half sibs), reliability of sires (>0.97) or sons (>0.72). In spite of a favorable genetic structure [linkage disequilibrium equal to 0.24 at 50 kb pairs], results showed low advantage of genomic evaluation. On the validation sample SF + FH + AA, the correlation between deregressed proofs and GBLUP or BayesC π predictions was 0.39, 0.37, 0.51 according to the different validation data sets, compared with 0.36, 0.33, 0.53 obtained with BLUP predictions. Correlations were much lower on the SF + FH sample. Research is pursued to understand this low advantage of genomic selection and to improve the methodology for genomic evaluation in this context, which is less favorable than dairy cattle breeding.

Key words: deregression, genomic selection, horse, jumping, SNP

© 2013 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2013.91:1076–1085
doi:10.2527/jas2012-5256

INTRODUCTION

Genomic selection (Meuwissen et al., 2001) has radically changed selection schemes in dairy cattle (Schaeffer, 2006; Wensch-Dorendorf et al., 2011) and is currently being studied in a number of species, such as pigs (Lillehammer et al., 2011), poultry (Preisinger, 2010; Chen et al., 2011; Wolc et al., 2011), and sheep

(Duchemin et al., 2012). In horse breeding, pedigree-based EBV are commonly developed in horse populations, especially within Europe, but still scarcely used by breeders as a tool for selection (Thorén Hellsten et al., 2006; Dubois and Ricard, 2007). However, from the time of the appearance of the first horse genome sequence (Wade et al., 2009), SNP genome markers have been used for marketing purposes in the trade of live animals (Hill et al., 2010, 2012). Genomic selection has potential interest in horse breeding: the cost of genotyping with the Equine SNP BeadChip assay is low compared with the value of the animal, the generation interval is long due to the late entrance in reproductive life of performer horses, the maternal way is underex-

¹The JUMPSNP program was funded by grants from the Institut National de la Recherche Agronomique (INRA, Paris, France), the Fonds Eperon (Paris, France), and the Institut Français du Cheval et de l'Équitation (IFCE, Saumur, France).

²Corresponding author: anne.ricard@toulouse.inra.fr

Received February 29, 2012.

Accepted November 21, 2012.

exploited, and other traits than performances are still difficult to improve in classical breeding schemes (Dubois et al., 2008).

Genomic evaluations often use as pseudo-phenotypes corrected means of progeny performances, like daughter yield deviations (**DYD**) in dairy species (VanRaden and Wiggans, 1991). In horse breeding, own performances are also available, and performances from other relatives (as half sibs) may play an important part in the EBV because the number of progeny remains low, even for stallions. Therefore, the simple procedures of VanRaden and Wiggans (1991) or Garrick et al. (2009) cannot be blindly used. The first step for genomic selection in horses is therefore to generate pseudo-phenotypes for genomic analysis when parental or own information is considered. This is the aim of this study. As an application, the reliability of genomic selection in the current breeding population of jumping horses in France was verified by cross-validation.

MATERIALS AND METHODS

Computation of Pseudo-Phenotypes

The first step for genomic selection in horses, and other species where the use of DYD is inappropriate, is to generate pseudo-phenotypes (unless the single step is used: Aguilar et al., 2010; Christensen and Lund, 2010). Deregressed EBV seem to be the right choice for the pseudo-phenotype to use all performances included in the estimation of these breeding values and to benefit from correction of fixed effects and preferential mating of stallions. More, use of deregressed EBV permits one to easily use multiple traits and nonlinear traits as single performance. To remain easy to use and general, we assume that the available data to compute pseudo-phenotypes (i.e., deregressed proofs) are EBV and their associated reliabilities.

Equivalent Models. The objective is to provide pseudo-phenotypes for genomic evaluation in such a way that these pseudo-phenotypes can be used in a simple genomic model, for instance this GBLUP:

$$\mathbf{y}_1^{(eq)} = \mathbf{1}\mu^{(eq)} + \mathbf{u}_1 + \mathbf{e}_1^{(eq)} \quad [1]$$

where the subscript $_1$ refers to the genotyped population; $\mathbf{y}_1^{(eq)}$ is the vector of pseudo-phenotypes, which summarizes all performances of ungenotyped relatives corrected for fixed effects; and $\text{Var}[\mathbf{e}_1^{(eq)}] = \mathbf{D}^{(eq)}\sigma_e^2$, where $\mathbf{D}^{(eq)}$ is a diagonal matrix with coefficients measuring the amount of information brought by these pseudo-phenotypes (i.e., weights), σ_e^2 is the residual variance

of the original pedigree-based genetic evaluation model, and (eq) stands for “equivalent” model. This model will be used with genomic data, where $\text{Var}(\mathbf{u}_1) = \mathbf{G}_{11}\sigma_u^2$, with \mathbf{G}_{11} the genomic relationship matrix between genotyped horses calculated from SNP genotypes. This was computed as

$$\mathbf{G}_{11} = \frac{\mathbf{W}\mathbf{W}'}{2 \sum_{j=1}^{4444} p_j (1-p_j)}$$

(VanRaden, 2008), with \mathbf{W} a centered incidence matrix of markers effects and p_j the allelic frequency of SNP j .

The pseudo-performances and weights in Eq. [1] must provide the same breeding values when using pedigree-based relationships for genotyped individuals as the full pedigree-based genetic evaluations, so that EBV $\hat{\mathbf{u}}_1$ must be the solution to

$$\begin{bmatrix} \mathbf{1}'\mathbf{D}^{(eq)-1}\mathbf{1} & \mathbf{1}'\mathbf{D}^{(eq)-1} \\ \mathbf{1}\mathbf{D}^{(eq)-1} & \mathbf{D}^{(eq)-1} + \lambda\mathbf{A}_{11}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu}^{(eq)} \\ \hat{\mathbf{u}}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{D}^{(eq)-1}\mathbf{y}_1^{(eq)} \\ \mathbf{D}^{(eq)-1}\mathbf{y}_1^{(eq)} \end{bmatrix}$$

with $\lambda = \frac{1-t}{h^2}$, and t is the repeatability of the performances and h^2 the heritability; and $\hat{\mathbf{u}}_1$ must also be the solution to

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix},$$

with $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, the original model used to calculate EBV where \mathbf{y} was the vector of raw performances of whole population, \mathbf{b} the vector of fixed effects, \mathbf{u} the vector of breeding values for all the population, \mathbf{X} and \mathbf{Z} incidence matrices.

The objective is to calculate $\mathbf{y}^{(eq)}$ and $\mathbf{D}^{(eq)}$ from these 2 expressions and available data: EBV estimates $\hat{\mathbf{u}}$ and their reliabilities, \mathbf{r} . The reduction of reliability due to the estimation of fixed effects was neglected. Thus, we focused on the second line of regular BLUP equations:

$$[\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1}][\hat{\mathbf{u}}] = [\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\hat{\mathbf{b}}]. \quad [2]$$

Partitioning the \mathbf{A} matrix according to genotyped ($_1$) and ungenotyped ($_2$) individuals gives

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \text{ and let } \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}$$

and $\mathbf{Z} = [\mathbf{Z}_1 \quad \mathbf{Z}_2]$.

The System [2] can be written as

$$\begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 + \lambda\mathbf{A}^{11} & \lambda\mathbf{A}^{12} \\ \lambda\mathbf{A}^{21} & \mathbf{Z}'_2\mathbf{Z}_2 + \lambda\mathbf{A}^{22} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ \mathbf{Z}'_2(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{bmatrix}$$

Absorbing $\hat{\mathbf{u}}_2$,

$$\begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 + \lambda\mathbf{A}^{11} - (\lambda\mathbf{A}^{12}) \\ (\mathbf{Z}'_2\mathbf{Z}_2 + \lambda\mathbf{A}^{22})^{-1}(\lambda\mathbf{A}^{21}) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - (\lambda\mathbf{A}^{12}) \\ (\mathbf{Z}'_2\mathbf{Z}_2 + \lambda\mathbf{A}^{22})^{-1}\mathbf{Z}'_2(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{bmatrix}$$

Let

$$\mathbf{M} = \begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 - (\lambda\mathbf{A}^{12})(\mathbf{Z}'_2\mathbf{Z}_2 + \lambda\mathbf{A}^{22})^{-1} \\ (\lambda\mathbf{A}^{21}) + \lambda(\mathbf{A}^{11} - \mathbf{A}_{11}^{-1}) \end{bmatrix},$$

and the system becomes

$$[\mathbf{M} + \lambda\mathbf{A}_{11}^{-1}]\begin{bmatrix} \hat{\mathbf{u}}_1 \\ \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - (\lambda\mathbf{A}^{12}) \\ (\mathbf{Z}'_2\mathbf{Z}_2 + \lambda\mathbf{A}^{22})^{-1}\mathbf{Z}'_2(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{bmatrix} \quad [3]$$

Let \mathbf{M}_d be a diagonal matrix giving the same solutions as \mathbf{M} in Eq. [3]. In principle, for animals with a reasonable reliability, this matrix is very similar to the diagonal of \mathbf{M} , and we will assume so. Setting Eq. [1] to be equivalent to Eq. [3] gives

$$\mathbf{D}^{(eq)-1} = \mathbf{M}_d$$

$$\begin{aligned} \mathbf{y}^{(eq)} &= \mathbf{M}_d^{-1} \begin{bmatrix} \mathbf{Z}'_1(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - (\lambda\mathbf{A}^{12}) \\ (\mathbf{Z}'_2\mathbf{Z}_2 + \lambda\mathbf{A}^{22})^{-1}\mathbf{Z}'_2(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{bmatrix} \\ &= \mathbf{M}_d^{-1} [\mathbf{M}_d + \lambda\mathbf{A}_{11}^{-1}] \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \end{bmatrix} \end{aligned}$$

Computation of Pseudo-Phenotypes and Weights.

So, \mathbf{M} (or more exactly its diagonal) has to be estimated from EBV and reliabilities (matrices in Eq. [2] are too large and, due to the nonlinearity of the rank evaluation in horses, not really sparse enough to be explicitly handled). From the reliabilities (\mathbf{r}) the different PEV can be computed, and therefore the diagonal of $(\mathbf{M} + \lambda\mathbf{A}_{11}^{-1})^{-1}$ can readily be formed. To obtain \mathbf{M} , one needs to infer which

part of the reliabilities is due to \mathbf{M} and which part to $\lambda\mathbf{A}_{11}^{-1}$, that is, to know for each animal the reliability which would have been obtained had it no relatives genotyped. The Harris and Johnson (1998) formulae enable us to separate the reliabilities according to different sources: parent average, the record of the animal, and progeny records. But their algorithm, which cycles through a complete pedigree, cannot be applied only in the genotyped population ($_1$) because not all founders are genotyped; in practice, relationships are very diverse and not only parent and progeny. On the other hand, the formulae would be easy to use if parents of the genotyped population were added to the system, and therefore genealogical paths between 2 (genotyped) relatives can be followed. So, the ungenotyped population ($_2$) was divided between the subpopulation ($_{2a}$) of ancestors of genotyped individuals ($_1$), and the remaining subpopulation ($_{2b}$). In that case, the original System [2] becomes:

$$\begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 + \lambda\mathbf{A}^{11} & \lambda\mathbf{A}^{12a} & \lambda\mathbf{A}^{12b} \\ \lambda\mathbf{A}^{2a1} & \mathbf{Z}'_{2a}\mathbf{Z}_{2a} + \lambda\mathbf{A}^{2a2a} & \lambda\mathbf{A}^{2a2b} \\ \lambda\mathbf{A}^{2b1} & \lambda\mathbf{A}^{2b2a} & \mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_{2a} \\ \hat{\mathbf{u}}_{2b} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ \mathbf{Z}'_{2a}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ \mathbf{Z}'_{2b}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{bmatrix}$$

And the absorption of Population ($_{2b}$) gives the System [4]:

$$\begin{bmatrix} \mathbf{Z}'_1 - \lambda\mathbf{A}^{12b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\lambda\mathbf{A}^{2b1} + \lambda\mathbf{A}^{11} & \lambda\mathbf{A}^{12a} - \lambda\mathbf{A}^{12b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\lambda\mathbf{A}^{2b2a} \\ \lambda\mathbf{A}^{2a1} - \lambda\mathbf{A}^{2a2b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\lambda\mathbf{A}^{2b1} & \mathbf{Z}'_{2a}\mathbf{Z}_{2a} - \lambda\mathbf{A}^{2a2b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\lambda\mathbf{A}^{2b2a} + \lambda\mathbf{A}^{2a2a} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_{2a} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - \lambda\mathbf{A}^{12b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\mathbf{Z}'_{2b}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ \mathbf{Z}'_{2a}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - \lambda\mathbf{A}^{2a2b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\mathbf{Z}'_{2b}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{bmatrix} \quad [4]$$

In that system, let's split the information due to relationships inside Population ($_{1+2a}$) and the remaining information. In fact, ($_{1+2a}$) form a complete pedigree, smaller than ($_{1+2a+2b}$). Let

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}^{11} & \mathbf{B}^{12a} \\ \mathbf{B}^{2a1} & \mathbf{B}^{2a2a} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12a} \\ \mathbf{A}_{2a1} & \mathbf{A}_{2a2a} \end{bmatrix}^{-1}$$

the inverse of relationship matrix in Population ($_1$) and ($_{2a}$); matrix \mathbf{B} is sparse and can be obtained using Henderson's rules) and let

$$\mathbf{K} = \begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 - \lambda\mathbf{A}^{12b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\lambda\mathbf{A}^{2b1} + \lambda(\mathbf{A}^{11} - \mathbf{B}^{11}) & \lambda(\mathbf{A}^{12a} - \mathbf{B}^{12a}) - \lambda\mathbf{A}^{12b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\lambda\mathbf{A}^{2b2a} \\ \lambda(\mathbf{A}^{2a1} - \mathbf{B}^{2a1}) - \lambda\mathbf{A}^{2a2b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\lambda\mathbf{A}^{2b1} & \mathbf{Z}'_{2a}\mathbf{Z}_{2a} - \lambda\mathbf{A}^{2a2b}(\mathbf{Z}'_{2b}\mathbf{Z}_{2b} + \lambda\mathbf{A}^{2b2b})^{-1}\lambda\mathbf{A}^{2b2a} + \lambda(\mathbf{A}^{2a2a} - \mathbf{B}^{2a2a}) \end{bmatrix}$$

The System [4] may be written as:

$$\begin{bmatrix} \mathbf{K}_{11} + \lambda \mathbf{B}^{11} & \mathbf{K}_{12a} + \lambda \mathbf{B}^{12a} \\ \mathbf{K}_{2a1} + \lambda \mathbf{B}^{2a1} & \mathbf{K}_{2a2a} + \lambda \mathbf{B}^{2a2a} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_{2a} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_{2a} \end{bmatrix}. \quad [5]$$

In these equations, the Harris and Johnson (1998, Appendix 1) formulae, reversed, enable us to compute the element of Δ , a diagonal matrix such as $[\Delta + \lambda \mathbf{B}]^{-1}$ gives the same diagonal coefficients as $[\mathbf{K} + \lambda \mathbf{B}]^{-1}$, which was not possible from Eq. [3]. So, Eq. [5] was replaced by

$$\begin{bmatrix} \Delta_{11} + \lambda \mathbf{B}^{11} & \lambda \mathbf{B}^{12a} \\ \lambda \mathbf{B}^{2a1} & \Delta_{2a2a} + \lambda \mathbf{B}^{2a2a} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_{2a} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_{2a} \end{bmatrix}. \quad [6]$$

The vector \mathbf{w} was easy to obtain from Δ , and the relationship matrix of the population of genotyped horses ($_1$) and their ancestors ($_{2a}$). This relationship matrix did not involve other horses.

Now, Δ and \mathbf{w} were known, and Eq. [6] was equivalent to

$$\begin{bmatrix} \Delta_{11} - \lambda \mathbf{B}^{12a} (\Delta_{2a2a} + \lambda \mathbf{B}^{2a2a})^{-1} \lambda \mathbf{B}^{2a1} + \lambda \mathbf{B}^{11} \\ \hat{\mathbf{u}}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 - \lambda \mathbf{B}^{12a} (\Delta_{2a2a} + \lambda \mathbf{B}^{2a2a})^{-1} \mathbf{w}_{2a} \\ \mathbf{w}_{2a} \end{bmatrix}.$$

Note that $\mathbf{A}_{11}^{-1} = \mathbf{B}^{11} - \mathbf{B}^{12a} (\mathbf{B}^{2a2a})^{-1} \mathbf{B}^{2a1}$, so that the system was:

$$\begin{bmatrix} \Delta_{11} - \lambda \mathbf{B}^{12a} ((\Delta_{2a2a} + \lambda \mathbf{B}^{2a2a})^{-1} + (\mathbf{B}^{2a2a})^{-1}) \lambda \mathbf{B}^{2a1} + \lambda \mathbf{A}_{11}^{-1} \\ \hat{\mathbf{u}}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 - \lambda \mathbf{B}^{12a} (\Delta_{2a2a} + \lambda \mathbf{B}^{2a2a})^{-1} \mathbf{w}_{2a} \\ \mathbf{w}_{2a} \end{bmatrix}$$

Off diagonal coefficients created by the absorption of equations of Population ($_{2a}$) were neglected and only diagonal coefficients of $\Delta_{11} - \lambda \mathbf{B}^{12a} ((\Delta_{2a2a} + \lambda \mathbf{B}^{2a2a})^{-1} + (\mathbf{B}^{2a2a})^{-1}) \lambda \mathbf{B}^{2a1}$ were retained. This corresponds to the this diagonal matrix \mathbf{M}_b sought after, then

$$\mathbf{D}^{(eq)-1} = \mathbf{M}_b,$$

and the pseudo-phenotypes were

$$\mathbf{y}^{(eq)} = \mathbf{M}_b^{-1} [\mathbf{w}_1 - \lambda \mathbf{B}^{12a} (\Delta_{2a2a} + \lambda \mathbf{B}^{2a2a})^{-1} \mathbf{w}_{2a}].$$

Application to French Jumping Data

Genotypic Data. The choice of the population to be genotyped was motivated by 2 criteria: the availability of a DNA sample, and the reliability of the estimated jumping breeding value of the stallion. In practice, only

around 850 stallions are currently alive and used to produce sport horses, and conserved DNA samples of dead stallions were not always available in sufficient quantity and quality. Putting together fresh blood samples and frozen semen of national stallions and privately owned stallions, as well as conserved DNA owned by the paternity validation laboratory, 908 horses were genotyped.

These horses were from these breeds: 71% Selle Français (SF), 17% foreign sport horses [FH, most of them Koninklijk Warmbloed Paardenstamboek Nederland (KWPN), Holsteiner, Belgian Warmblood, Hannoverian, Oldenburg, and Cheval de Sport Belge], and 13% Anglo-Arabian (AA). Most horses were stallions; only 17 mares were included. They were mainly born from 1989 to 2005, equally distributed among these years (i.e., about 46 horses per year). A few horses (14%) were born earlier, up to 1974. Most horses had own performances in competition (92%). A majority of stallions were, in addition, sires of competing horses in jumping (61%). In the genotyped individuals, there were 78% of the SF stallions in reproductive activity in 2009, as were 45% of the FH stallions, and 65% of the AA stallions. Therefore, these stallions i) are representative of the breeding population and ii) do have a sizeable amount of information. Thus, it would be difficult to enlarge this database.

Among these 908 horses, 127 were parents or grandparents of some other genotyped horse; conversely, 336 horses had their sire genotyped. There were 342 genotyped half-sib families, with an average size of 2.7. The 342 families included 204 families composed of a single couple father-son, 82 families with 3 half sibs or more, and 17 families of 10 half sibs or more (up to 39).

Markers Information. Horse genotyping was performed using the Illumina Equine SNP50 BeadChip assay at Labogena (Jouy en Josas, France), according to the manufacturer's standard procedures. This array includes 54,602 SNP evenly distributed throughout the genome. Poor quality markers were discarded based on 3 criteria: markers genotyped in less than 80% of the samples (call frequency < 80%), or having a minor allele frequency (MAF) under 5%, or strongly deviating from Hardy-Weinberg equilibrium (P -value < 10^{-8}). The X chromosome was not included. Finally, 44,444 SNP were retained on the 31 autosomal chromosomes. The mean distance between adjacent SNP was 50,256 bp (~0.05 cM); about half adjacent SNP (48%) were closer than 25,000 bp. The MAF was uniformly distributed between 5 and 50%.

Genetic and Genomic Evaluations. Two genomic models were used. The first is the GBLUP as in Eq. [1], detailed in the computation of pseudo-phenotypes. The second is the BayesC π (Habier et al., 2011), where the phenotypes $\mathbf{y}_1^{(eq)}$ are a function of marker effects

$$\mathbf{y}_1^{(eq)} = 1\boldsymbol{\mu}^{(eq)} \sum_{j=1}^{4444} z_{ij} a_j \delta_j + \mathbf{e}_1^{(eq)},$$

where z_{ij} is an incidence coefficient, a_j is the effect of marker j , and δ_j is an indicator variable stating whether the marker j has any effect (1) or not (0), and $\boldsymbol{\mu}^{(eq)}$ and $\mathbf{e}_1^{(eq)}$ are as previously noted. The distribution for $\boldsymbol{\delta} = (\delta_1 \dots \delta_{4444})$ is a multivariate Bernoulli such as $P(\delta_1 = 1|\pi) = \pi$, where π is the probability that the marker has an effect on the phenotype; π was fixed to 0.09, given our previous experiences in dairy sheep (Duchemin et al., 2012) and cattle (Colombani et al., 2013) which showed good accuracies for this value and a general robustness of BayesCPI to the exact value of π . The prior distribution of marker effects was normal with a common variance after an inverted χ^2 distribution. Therefore, the marker variance σ_a^2 is common to all loci. For details of the model, see Habier et al. (2011). Breeding values were obtained through GS3 software developed by Legarra et al. (2010) and were called BAYESCPI. The posterior distribution was computed using a full Monte Carlo Markov Chain of 50,000 iterations, with a burn-in of 200 iterations.

To compare genomic evaluation to traditional evaluation, the Model [1] was also applied with genealogical data assuming $\text{Var}(\mathbf{u}_1) = \mathbf{A}_{11}\sigma_u^2$, with \mathbf{A}_{11} the relationship matrix between genotyped horses. This evaluation was called BLUP.

Estimated Breeding Values, Genetic Parameters, and Genealogy. The original EBV were based on a multiple-trait BLUP animal model with 2 highly (0.90) genetically correlated criteria: logarithm of annual sum of points and an underlying variable responsible for ranking in each event (Ricard, 1997). The multiple trait model is used only to deal with special cases in phenotypes (e.g., horses with a low number of events). Because the 2 traits are highly correlated, in the remaining, the EBV will be considered as calculated from a single, repeatable composite trait. The original EBV involved 242,620 horses with 1,006,414 annual performances and more than 14 million events, and included 422,522 relatives.

To compute $\mathbf{y}^{(eq)}$, $\mathbf{D}^{(eq)}$, and \mathbf{A}_{11} , the genetic relationship matrix used for BLUP, the pedigree of genotyped horses up to all known ancestors (whatever the breed) born after 1945 was used. The total file represented 6562 horses.

The heritability of the trait was considered as fixed, whatever the models to suppress the uncertainty in the estimations, due to the rather small data size. Heritability of jumping results in competition was 0.25 and repeatability of the trait between years 0.45, which gave $\lambda = 2.75$. For BAYESCPI, the same genomic variance was obtained combining $\pi = 9\%$, $\sigma_a^2 = 0.00014$, and $\sigma_a^2 = 0.55$.

Training and Validation Data Set. Usually, comparisons of genomic and pedigree-based evaluation use 1 training and 1 validation data set which mimic the true accumulation of information across time (Hayes et al., 2009b). The predicted EBV of young bulls without progeny in the training data set are compared with the realized DYD obtained later. In absence of strong selection, and with the equivalent model developed here, any kind of splitting of validation and training data sets could in principle be used, because information from relationships in the genotyped sample was separated from all other information. However, as pointed out by Hayes and Goddard (2008), Goddard (2009), and Habier et al. (2010), the accuracies of GEBV depend on relationships between training and validation data set and reliability of breeding values in the training data set. Moreover, the correlation between EBV and pseudo-phenotypes in the validation data set depends on the accuracy of breeding values used to compute pseudo-phenotypes. Therefore, we constructed a validation data set which included stallions meeting the following conditions: with sire genotyped in training data set, with weight > 3 , and meaning reliability > 0.52 , to assure minimum links between training and validation data and minimum information about genetic value of the stallions in the validation data set. Then we tried different structures of validation data sets to highlight the genetic differences inside the population detectable by genomics. We used 3 different validation data sets, including stallions: issued from families with at least 3 half sibs (Data Set 1); with reliability of the sire of the stallion over 0.97 in training (weight > 100 ; Data Set 2); and with reliability of the stallion over 0.72 (weight > 7 ; Data Set 3). In each of these 3 cases, the training data set was defined as all genotyped stallions not included in the validation data set.

Two estimations were performed for each data set: the first with all breeds, and the second with only SF and FH in both training and validation data sets. The comparison was made with correlation and regression between BLUP, GBLUP, BAYESCPI estimated EBV, and pseudo-phenotypes.

RESULTS

Population Structure

Figure 1 shows linkage disequilibrium (LD) measured by average r^2 between syntenic SNP pairs against map distance up to 1 megabase according to the breed. All available pairs of SNP were used and grouped by steps of 5 kb. Linkage disequilibrium was 0.280 for AA, 0.228 for FH, and 0.243 for SF at the mean distance (50 kb) between adjacent markers. There were 216 monomorphic SNP in AA, and 1 in SF and FH. The corre-

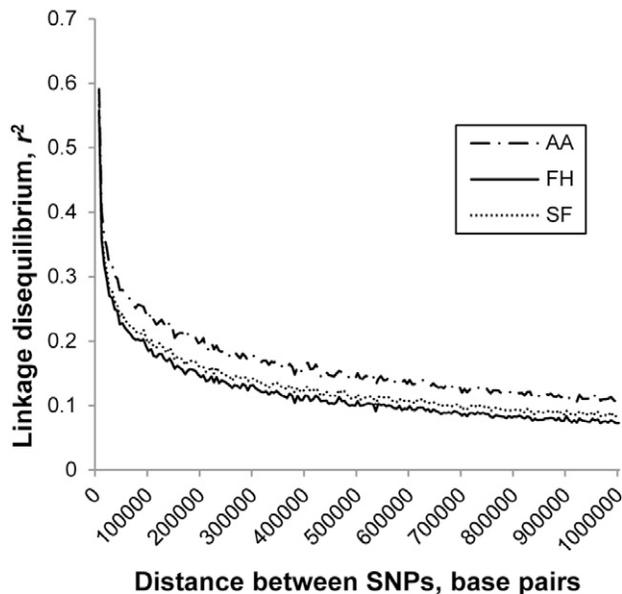


Figure 1. Linkage disequilibrium (r^2) in Selle Français horses (SF), foreign sport horses (FH), and Anglo-Arab horses (AA) assessed by using 908 samples from EquineSNP50 genotypes (each point is the mean of all SNP pairs over 5 kb pairs).

lation between r^2 measured on adjacent SNP (44,001 pairs) in each breed was 0.956 between SF and FH, 0.918 between SF and AA and 0.907 between FH and AA. This implies a very strong concordance of phase across breeds and therefore a low divergence.

Figure 2 shows the 2 first principal components of principal component analysis of the genomic relationship matrix for all analyzed animals. The AA horses form a very different group from the others, whereas SF and FH tend to mixed together. For SF and FH, rules for assigning crossed horses to a breed have varied with time. By adding the percentage of thoroughbred in SF horses in the plot, 4 main axes were distinguished: FH, AA, Thoroughbred, and French saddle horse.

Pseudo-Performances and Weights

The equivalent performance was expressed in units of phenotypic SD. The distribution was nearly normal, the mean was 1.374 (SD = 0.743, with minimum of -0.898, and maximum of 3.949). The weights had mean 16.869 (± 31.563 , with minimum of 1.150, and maximum of 341.609). These weights corresponded to the reliability of breeding values, corrected as if relationships were unknown in the genotyped population. The average reliability was 0.67 (SD = 0.19, minimum of 0.30, and maximum of 0.99).

Validation

The results of cross validation are shown in Table 1. The correlation between pseudo-phenotypes and genomic evaluation (GBLUP, BAYESCPI) is greater than

Table 1. Correlation and regression between pseudo-phenotypes for show jumping and genetic evaluation from genealogy (BLUP), or from SNP using genomic BLUP method (GBLUP) or BayesC π method (BayesCPI) in 3 different validation data sets.

Genetic evaluation	Correlation		Regression	
	All breeds	SF + FH ¹	All breeds	SF + FH
Validation set 1 ²				
<i>N</i>	103	84	103	84
BLUP	0.36	0.28	0.99	0.87
GBLUP	0.39	0.30	0.77	0.72
BayesCPI	0.39	0.29	0.74	0.69
Validation set 2 ³				
<i>N</i>	98	89	98	89
BLUP	0.33	0.30	0.97	0.89
GBLUP	0.37	0.34	0.71	0.73
BayesCPI	0.37	0.33	0.69	0.70
Validation set 3 ⁴				
<i>N</i>	91	76	91	76
BLUP	0.53	0.46	0.86	0.80
GBLUP	0.51	0.47	0.64	0.68
BayesCPI	0.51	0.47	0.62	0.66

¹SF = Selle Français, FH = foreign sport horses.

²Horses with their sire genotyped in training data set, minimum reliability = 0.52, issued from families of at least 3 half sibs.

³Horses with their sire genotyped in training data set and minimum reliability = 0.97, minimum reliability = 0.52.

⁴Horses with their sire genotyped in training data, minimum reliability = 0.72.

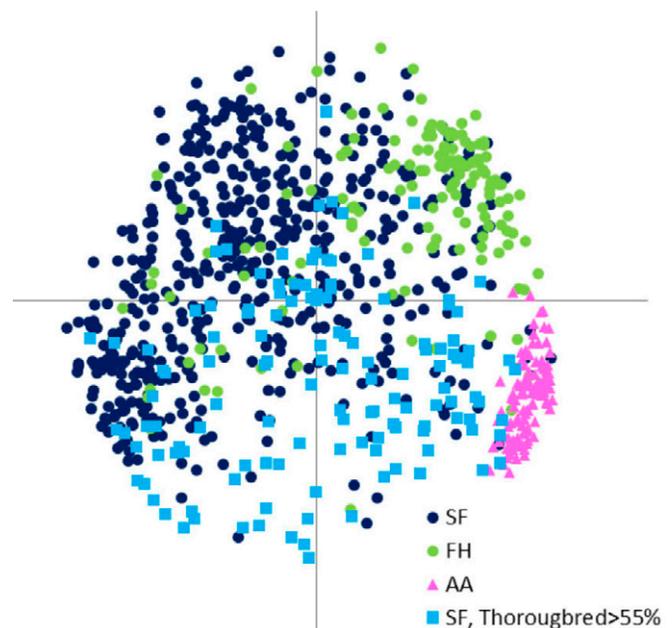


Figure 2. Plot of the 908 genotyped horses in function of the first 2 principal components of the genomic matrix according to the breed (SF = Selle Français, FH = foreign sport horses, AA = Anglo-Arab). See online version for figure in color.

the correlation with pedigree genetic evaluation (BLUP), except for the Validation Set 3 multibreed (validation stallions chosen based on their reliability). But this superiority was low, that is, from 0.01 to 0.04. The regression coefficients were near 1 for the genetic evaluation, but much lower for the genomic 1, about 0.70.

DISCUSSION

What is Improved with the Use of Deregressed EBV?

Compared with DYD, the use of deregressed proofs has the same objective: to summarize performances from ungenotyped progeny to reach the genetic value of the genotyped horse, and getting rid at the same time of confounding effects (sex, age, year, assortative mating). Of course, deregressed proofs, as opposed to DYD, can include and properly weight performances from relatives other than progeny (e.g., own performances and half-sib performances). In the sample used, almost all stallions had own performances but only 61% had progeny, allowing a larger data set than if using only progeny-tested sires and more information for each horse to be included in the analysis. Moreover, the use of deregression clearly separates the information inside and outside of the genotyped sample meaning that, if a female created a link inside the genotyped sample (e.g., as mare of stallion) and is also used as mate of other stallions, these 2 functions are distinguished. This is not the case when using DYD, where the full EBV of the dam is included in the correction of the performances of daughters, whatever the links with other bulls. This split between covariances inside and outside the genotyped sample gave formulae (not shown) which showed a clear relationship of our deregressions with the single step method (Aguilar et al., 2010, Christensen and Lund, 2010) as introducing the same deviation of the relationship matrix from the expected one based on pedigree. However, the computations needed in our deregression assume matrices (e.g., \mathbf{M}) close to diagonal and, therefore, are valid for animals with relatively large information, which was the case of our sample.

How to Use Deregressed Proofs?

Garrick et al. (2009) proposed 2 notions to correctly use deregressed proofs: i) removing parent average effects, and ii) weighting deregressed information according to the percentage of genetic variance explained by SNP. For removing parental effects (PA), their arguments are based on the possibility of the effect of a major gene. In that case, the EBV of the genotyped individual will be shrunken toward the parent average whatever the allele inherited. Of course, regular BLUP does not take into account specifically such an inheritance. But if there is a

large gene segregating, the same will be true for all the process involving generation of DYD: wrong estimation of fixed effects and excessive shrinkage of the mates of the bull (thus, the accuracy of the DYD is overestimated). Anyhow, major gene or not, DYD will be unbiased [e.g., $E(\overline{\text{DYD}}) = E(\text{DYD})$], and deregressed proofs will be unbiased as well. The argument of Garrick et al. (2009) seem to rely more on the difference between expected and realized transmission from parents to progeny, which cannot be accounted for in regular pedigree-based evaluations (but it is, for instance, in Fernando and Grossman, 1989). It has to be kept in mind that PA includes all information from relatives which are not progeny (e.g., half- and full-sibs). Therefore, the option is to throw away parental information (which might be serious for certain species or traits) or to include it, knowing that there is a risk that the polygenic model might be wrong, but the practice shows that this is rarely the case. Anyhow, this risk occurs everywhere in genomic or pedigree-based evaluations. And at any rate, PA are explicitly included in the process of estimating DYD (unless finite locus models are used for evaluation). Thus, it is not because we do not add PA to DYD that the latter are free from troubles. One could equally argue that neither DYD nor deregressed proofs are correct, because they rely in estimates of the breeding value of parents and mates which are shrunken, and therefore one should work with crude phenotypes (e.g., as in a sire model). Of course we know that this is incorrect when selection exists. Therefore, in our opinion, there is no really valid argument to reject the use of PA in either pedigree-based or genomic-based models, besides the good computational properties of DYD and their ease of interpretation (i.e., the phenotype of a bull, had it been a cow).

Composite Performances against Raw Performances?

On the other hand, we need to point out that, whatever the performances of relatives used to create the pseudo-performance of a genotyped individual (DYD or deregressed proofs as here), this information is obtained using genealogical rules, not any marker-based (genomic) rule, due to the lack of other information (e.g., the probabilities of transmission in Fernando and Grossman, 1989). So, in a sense, creating pseudo-performances ought to favor the genealogical evaluation compared with the genomic evaluation. So, what can be done? In our opinion, the correct strategy is to distinguish information inside and outside the genotyped sample. This is what we did in this work. The pseudo-phenotypes were thus more correct than DYD, because i) we used all available information, and ii) the amount of information from progeny was correctly separated into paternal and maternal path for genetic and genomic evaluation, so that none of it had advantages. Nevertheless,

this was not the limiting aspect of the study. And better results obtained in dairy cattle with DYD than in jumping horses with deregressed proofs must be explained in another way. The theoretical model used in GBLUP is not different from that used in classical breeding evaluation, except that the covariance between genetic values are different (one from genealogy, the other from markers). Our deregression method is equivalent to an acyclic graph (i.e., there is no notion of “time”). Therefore, in principle, any partition in training and validation is valid for crossvalidation (with the proviso that selection does not occur or it is of low intensity). This is not the case with DYD where, for instance, performances of granddaughters must not appear in the genetic evaluation of the training data set. Therefore, cross validation in dairy cattle is always performed including a notion of “time” in the choice of training (reduced data set) and a validation sample (full data set), with 4 more years of cumulated data (Mantysaari et al., 2010). By doing so, as a consequence, selection of sires is also taken into account as almost 1 generation includes testing and service bulls. In our sample, males with performances not taken for reproduction were not genotyped and never included in the analysis. Thus, the whole sample was selected, with no genomic information indicating the selection process. This may explain partly the results obtained in jumping horses, even with deregressed proofs.

Results on Horse Data: What Could Be Expected for the Efficiency of Genomic Selection Knowing the Population Structure?

The level of LD in the French horse population was comparable to other species where genomic selection was efficient. McKay et al. (2007) studied LD in 5 bovine breeds and found mean r^2 equal to 0.50 at 5 kb and 0.22 at 199 kb, so slightly greater than SF measurements (respectively, 0.47 and 0.20). The situation in sheep was more unfavorable, with r^2 between 0.12 and 0.19 at 50 kb (Kemper et al., 2011) against 0.24 for the same distance in SF. This relatively high LD is in principle favorable to genomic selection, as the SNP used might be linked to genes. On the other hand, the high LD on a long distance could have been unfavorable due to lack of accuracy in the localization of genes of interest and misattribution of such genes to correct horses.

The analysis of the genomic relationship matrix **G** revealed a clear difference between AA horses and the other sport horses. This is unfavorable for genomic evaluation (Hayes et al., 2009a) because only weak links between these 2 populations exist; therefore, genes spotted through SNP in 1 population are not useful for another population. But, because there was a difference for jumping abilities between AA and SF, the genomic evaluation tended to estimate the difference between breeds, rather

than differences between horses inside breeds. One choice would have been a model with the fixed effect of the breed, but it was not possible because the number of AA horses in the validation set was too small (from 9 to 15, depending on data set). So, we preferred to restrict the analysis to SF + FH, closer breeds. However, at first sight, the high correlation between LD observed for the same pairs of SNP in the different breeds could have been a favorable situation for the multibreed purpose.

Results on Horse Data: Is Finally the Genomic Selection Reliable for Jumping Horses?

Finally, the reliability of genomic selection on this data set for jumping horses was very low compared with what has been obtained in other species. The problem was not the trait studied, *jumping ability*, as the results of cross validation for BLUP were as good as, for example, in dairy cattle, considering the differences in reliabilities of the 2 pseudo-phenotypes (Hayes et al., 2009b; Olson et al., 2011). And the slope of the regression in that case was 1. Use of jumping performance in BLUP animal model was validated a long time ago in France (Tavernier, 1994), showing correlations between performances and genetic evaluation at birth between 0.31 and 0.36 (according to age at performance) and regression slopes between 0.80 and 1.14. The problem was the low difference between genomic (whatever the method) and genetic evaluations. Several factors already highlighted in literature may explain these results. First of all, the small sample size (Luan et al., 2009; VanRaden et al., 2009), but also the low level of relationships between training and validation samples (Habier et al., 2007, 2010; Goddard et al., 2011), the small accuracy of pseudo-performances (Goddard, 2009; Hayes et al., 2009b), and finally, the use of a multibreed sample (Hayes et al., 2009a).

Greater correlations are expected whenever sires of validation data will have more progeny, producing perhaps better results for genomic evaluations. But, in fact, the data structure that we used is similar to what will exist in real life, except that “old” information will cumulate throughout the years; but “old” information is not very useful for prediction of young born.

In practice, for selection of the youngest candidates, the pedigree-based BLUP, as well as the genomic evaluation, will have lower accuracies than those observed in this work. The reason is that both evaluations included information from pedigree which may not always be available in reality at the first stage of selection (but that was available in the cross validation). For example, grandchildren from other sons of the sire of the candidate might not yet be born when the candidate would be selected; however, they have been included in the evaluation of the sire in this cross validation. But this effect should remain low because the validation data set

was mostly built with young sires. This effect influenced both pedigree and genomic evaluation.

The performance of genomic selection in dairy cattle is empirically rather well understood. This is so because of the large number of studies and the excellent structure of the data set (accurate DYD, neat cut-offs between training, and validation samples). However, this is not the case for other kinds of populations, where there is not much experience or data are more complicated. Goddard et al. (2011) have outlined methods to estimate “a priori” accuracy of genomic selection. These methods consider existing LD in the population (which is actually a measure of realized historical co-ancestry), as well as the relative weights of genomic and pedigree relationships. It is our intention to extend this line of research to estimate whether our results make sense or not, and under which condition genomic selection would be promising.

In addition, our pseudo-phenotypes rely on assumptions of the infinitesimal model; therefore, it is likely that they favor this same infinitesimal model. It is not the case in dairy cattle, where DYD for production traits are extremely accurate. This hypothesis remains as well to be investigated.

Conclusion

In this paper, accuracy of genomic evaluation for show jumping was obtained from an almost exhaustive sample of living stallions in France. To do so, a specific deregression procedure was necessary to take into account own performances and performances of various relatives (not only progeny) outside the genotyped sample. This deregression procedure was easy to implement from official EBV, reliabilities, and pedigrees. Unfortunately, the accuracy of genomic evaluation, measured by cross-validation on different validation samples, was not sufficiently greater, to suggest at present an application in breeding plans for jumping horses. But this conclusion is only related to accuracy and potential benefits from greater selection intensity; shorter generation interval and lower inbreeding in the long term have also to be taken into account when planning genomic selection in horses. Research will be pursued to improve this result.

LITERATURE CITED

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.
- Chen, C. Y., I. Misztal, I. Aguilar, S. Tsuruta, T. H. E. Meuwissen, S. E. Aggrey, T. Wing, and W. M. Muir. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. *J. Anim. Sci.* 89:23–28.
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.
- Colombani, C., A. Legarra, S. Fritz, F. Guillaume, P. Croiseau, V. Ducrocq, and C. Robert-Granie. 2013. Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesCpi methods for genomic selection in French Holstein and Montbeliarde breeds. *J. Dairy Sci.* 96:575–591.
- Dubois, C., E. Manfredi, and A. Ricard. 2008. Optimization of breeding schemes for sport horses. *Livest. Sci.* 118:99–112.
- Dubois, C., and A. Ricard. 2007. Efficiency of past selection of the French Sport Horse: Selle Français breed and suggestions for the future. *Livest. Sci.* 112:161–171.
- Duchemin, S. I., C. Colombani, A. Legarra, G. Baloche, H. Larroque, J. M. Astruc, F. Barillet, C. Robert-Granie, and E. Manfredi. 2012. Genomic selection in the French Lacaune dairy sheep breed. *J. Dairy Sci.* 95:2723–2733.
- Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* 21:467–477.
- Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41:55.
- Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica (The Hague)* 136:245–257.
- Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* 128:409–421.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397.
- Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186.
- Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 42:5.
- Harris, B., and D. Johnson. 1998. Approximate reliability of genetic evaluations under an animal model. *J. Dairy Sci.* 81:2723–2728.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009b. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443.
- Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and M. E. Goddard. 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet. Sel. Evol.* 41:51.
- Hayes, B. J., and M. E. Goddard. 2008. Technical note: Prediction of breeding values using marker-derived relationship matrices. *J. Anim. Sci.* 86:2089–2092.
- Hill, E., J. S. Bolger, D. Ryan, and D. MacHugh. 2012. Equinome. www.equinome.com/index.html. (Accessed 4 March 2013.)
- Hill, E. W., J. J. Gu, S. S. Eivers, R. G. Fonseca, B. A. McGivney, P. Govindarajan, N. Orr, L. M. Katz, and D. E. MacHugh. 2010. A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS ONE* 5(1):e8645.
- Kemper, K. E., D. L. Emery, S. C. Bishop, H. Oddy, B. J. Hayes, S. Dominik, J. M. Henshall, and M. E. Goddard. 2011. The distribution of SNP marker effects for faecal worm egg count in sheep, and the feasibility of using these markers to predict genetic merit for resistance to worm infections. *Genet. Res.* 93:203–219.
- Legarra, A., A. Ricard, and O. Filangi. 2010. GS3-Genomic selection, Gibbs sampling, Gauss Seidel and BayesCπ. <http://snp.toulouse.inra.fr/~alegarra/>. (Accessed 4 March 2013.)
- Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011. Genomic selection for maternal traits in pigs. *J. Anim. Sci.* 89:3908–3916.

- Luan, T., J. A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T. H. E. Meuwissen. 2009. The accuracy of genomic selection in norwegian red cattle assessed by cross-validation. *Genetics* 183:1119–1126.
- Mantysaari, E. A., Z. Liu, and P. M. VanRaden. 2010. Interbull validation test for genomic evaluations. *Bull. 41. Interbull*, Uppsala, Sweden.
- McKay, S. D., R. D. Schnabel, B. M. Murdoch, L. K. Matukumalli, J. Aerts, W. Coppieters, D. Crews, E. Dias, C. A. Gill, C. Gao, H. Mannen, P. Stothard, Z. Q. Wang, C. P. Van Tassell, J. L. Williams, J. F. Taylor, and S. S. Moore. 2007. Whole genome linkage disequilibrium maps in cattle. *BMC Genet.* 8:74.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Olson, K. M., P. M. VanRaden, M. E. Tooker, and T. A. Cooper. 2011. Differences among methods to validate genomic evaluations for dairy cattle. *J. Dairy Sci.* 94:2613–2620.
- Preisinger, R. 2010. Genomic selection in poultry. *Zuchtungskunde* 82:9–13.
- Ricard, A. 1997. Breeding evaluations and breeding programs in France. *In* 48th Ann. Meet. EAAP, Vienna, Austria. p. 374.
- Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218–223.
- Tavernier, A. 1994. Special problems in genetic evaluation of performance traits in horse. *In* 5th world congress on genetics applied to livestock production, Guelph, ON, Canada. p. 450–457.
- Thorén Hellsten, E., A. Viklund, E. P. C. Koenen, A. Ricard, E. Bruns, and J. Philipsson. 2006. Review of genetic parameters estimated at stallion and young horse performance tests and their correlations with later results in dressage and show-jumping competition. *Livest. Sci.* 103:1–12.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24.
- VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal-model information. *J. Dairy Sci.* 74:2737–2746.
- Wade, C. M., E. Giuotto, S. Sigurdsson, M. Zoli, S. Gnerre, F. Imsland, T. L. Lear, D. L. Adelson, E. Bailey, R. R. Bellone, H. Blocker, O. Distl, R. C. Edgar, M. Garber, T. Leeb, E. Mauceli, J. N. MacLeod, M. C. T. Penedo, J. M. Raison, T. Sharpe, J. Vogel, L. Andersson, D. F. Antczak, T. Biagi, M. M. Binns, B. P. Chowdhary, S. J. Coleman, G. Della Valle, S. Fryc, G. Guerin, T. Hasegawa, E. W. Hill, J. Jurka, A. Kiialainen, G. Lindgren, J. Liu, E. Magnani, J. R. Mickelson, J. Murray, S. G. Nergadze, R. Onofrio, S. Pedroni, M. F. Piras, T. Raudsepp, M. Rocchi, K. H. Roed, O. A. Ryder, S. Searle, L. Skow, J. E. Swinburne, A. C. Syvanen, T. Tozaki, S. J. Valberg, M. Vaudin, J. R. White, M. C. Zody, Broad Inst. Genome Sequencing Platform, and T. Broad Inst Whole Genome Assembly Team, E. S. Lander, and K. Lindblad-Toh. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326:865–867.
- Wensch-Dorendorf, M., T. Yin, H. H. Swalve, and S. König. 2011. Optimal strategies for the use of genomic selection in dairy cattle breeding programs. *J. Dairy Sci.* 94:4140–4151.
- Wolc, A., C. Stricker, J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, S. J. Lamont, and J. C. M. Dekkers. 2011. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genet. Sel. Evol.* 43:5.

Supplementary Material

Supplementary material can be found at:

<http://www.journalofanimalscience.org/content/suppl/2013/04/03/jas.2012-5256.DC1.html>

References

This article cites 33 articles, 7 of which you can access for free at:

<http://www.journalofanimalscience.org/content/91/3/1076#BIBL>