Journal of Central European Agriculture, 2014, 15(2), p.33-40 DOI: 10.5513/JCEA01/15.2.1447

The influence of the genomic data on the showjumping horses performance evaluation

Vplyv genomických dát na hodnotenie výkonnosti parkúrových koní

Zuzana SCHUBERTOVÁ*, Ivan PAVLÍK and Juraj CANDRÁK

Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovak Republic, tel. +421-37-641 4289, xschubertova@is.uniag.sk * correspondence

Abstract

The aim of this study was to find out the influence of the genomic data on the showjumping horses performance evaluation for the purpose of genetic evaluation and subsequent estimation of breeding values of sport horses in Slovakia. For the evaluation of showjumping performance data of Global Champions Tour (season 2012) was used. The 4,729 individual competitions were recorded. The fourteen SNPs (single nucleotide polymorphisms) linked with showjumping performance were used for the simulation analyse. To explain the variability of horses ranking we used some different models. In the first model without SNP effect we are able to describe 53 % of performance (coefficient of determination R^2 was 0.53). In linear model with some effects in form of quadratic regression we obtained R^2 0.60. Highest R^2 0.61 we obtained in the third model where one SNP effect was taken into account. Coefficient of determination $R^2 = 0.63$ we obtained in the model where we used fourteen simulated SNP effects. It is possible to use simulated SNP data for evaluation of performance of showjumping horses. We observed increase of R^2 in dependence on number of involved SNPs in the model. In case of fourteen simulated SNP effects used in the linear model, increase of R^2 was 0.03 compared to the model without SNP effect.

Keywords: genomics, horses, showjumping performance, SNP

Abstrakt

Cieľom štúdie bolo zistiť vplyv genomických dát na hodnotenie výkonnosti parkúrových koní za účelom genetického zhodnotenia s následným odhadom plemenných hodnôt. Pre posúdenie výkonnosti parkúrových koní sme použili údaje zo súťaže Global Champions Tour (sezóna 2012). Zaznamenaných bolo 4 729 individuálnych súťaží. Simulovaných bolo 14 SNP, ktoré sú asociované s výkonnosťou parkúrových koní. Na vysvetlenie výkonnosti koní sme použili viacero lineárnych modelov, pričom sme hodnotili znak umiestnenie koní v súťažiach. V prvom modeli, v ktorom sme nezohľadnili SNP efekt, sme zistili koeficient determinácie R^2 0,53. V modeli, kde boli niektoré efekty vo forme kvadratickej regresie R^2 bol 0,60. Vyšší R^2 sme zistili v treťom modeli, kde sme zohľadnili aj jeden SNP efekt. V prípade použitia 14 SNP efektov, koeficient determinácie R^2 bol 0,63. SNP dáta je možné použiť pri odhade výkonnosti parkúrových koní. Zaznamenali sme zvýšenie R^2 v závislosti od množstva zapojených SNP efektov v modeli. V prípade zapojenia 14 SNP efektov v lineárnom modeli R^2 bol vyšší o 0,03 v porovnaní s lineárnym modelom bez zohľadnenia SNP efektu.

Kľúčové slová: genomika, kone, parkúrové skákanie, SNP

Detailed abstract in native language

Cieľom štúdie bolo zistiť vplyv genomických dát na hodnotenie výkonnosti parkúrových koní za účelom genetického hodnotenia s následným odhadom plemenných hodnôt. Pre posúdenie výkonnosti parkúrových koní sme použili údaje zo súťaže Global Champions Tour (sezóna 2012). Výška parkúru pri hodnotených súťažiach bola 110, 115, 120, 125, 130, 135, 140, 145, 150, 155 a 160 centimetrov (CSI1*, CSI2*, CSI5* and Special Invitational). Zaznamenaných bolo 4 729 individuálnych súťaží. Vek hodnotených koní (1 021) bol od 6 rokov do 20 rokov. Z hodnotenia sme vylúčili tie kone, ktoré boli diskvalifikované, nemali zaznamenaný čas alebo zisk trestných bodov v súťaži. Simulovaných bolo 14 SNP, ktoré sú asociované s výkonnosťou parkúrových koní. Na vysvetlenie výkonnosti koní sme použili viacero lineárnych modelov, pričom sme hodnotili znak umiestnenie koní v súťažiach. V prvom modeli, v ktorom sme nezohľadnili SNP efekt, sme zistili koeficient determinácie R^2 0,53. Všetky použité efekty boli vysoko štatisticky preukazné okrem efektu súťaž. V modeli, kde boli niektoré efekty vo forme kvadratickej regresie (vek, zisk trestných bodov, čas) R^2 bol 0,60. Všetky efekty boli vysoko štatisticky preukazné okrem efektu vek vo forme kvadratickej regresie. Vyšší R^2 sme zistili v treťom modeli, kde sme zohľadnili aj jeden SNP efekt. Všetky efekty boli vysoko štatisticky preukazné okrem efektu vek vo forme kvadratickej regresie, SNP efekt nebol štatisticky preukazný. V prípade použitia 14 SNP efektov, koeficient determinácie R^2 bol 0,63. Efekt vek vo forme kvadratickej regresie a SNP 11 efekt neboli štatisticky preukazné. Všetky ostatné efekty boli vysoko štatisticky preukazné alebo štatisticky preukazné. SNP dáta je možné použiť pri odhade výkonnosti parkúrových koní. Zaznamenali sme zvýšenie R^2 v závislosti od množstva zapojených SNP efektov v modeli. V prípade zapojenia 14 SNP efektov v lineárnom modeli R^2 bol vyšší o 0,03 v porovnaní s lineárnym modelom bez zohľadnenia SNP efektu.

Kľúčové slová: genomika, kone, parkúrové skákanie, SNP

Introduction

Plant and animal breeding programs are being transformed by the use of genomic data, which is becoming widely available and cost-effective to predict genetic merit. A large number of genomic prediction studies have been published using both simulated and real data (Daetwyler, et al., 2013). The availability of many thousands of singlenucleotide polymorphisms (SNPs) spread across the genome for different livestock species opens up possibilities to include genomewide marker information in prediction of total breeding values, to perform genomic selection (Calus, et al., 2008). Genomic selection can increase the rates of genetic gain through increased accuracy of estimated breeding values, reduction of generation intervals, and better utilization of available genetic resources through genome-guided mate selection (Sonesson, et al., 2010; Schierenbeck, et al., 2011). Simulation of genomic data is useful because it allows rapid replicated testing of a wide range of hypotheses at low cost, for example, the initial feasibility of genomic selection or impact of reference population size. Simulated data allows the researcher to explore important aspects such as the genetic architecture of the traits, number of markers used for analysis, degree of relatedness between the training and prediction population, and offers the possibility of evaluating some sources of variability, such as drift, which cannot be assessed with real data. Simulation of genomic data to the evaluation of animal performance and to the evaluation of breeding values has become widely used in some livestock industries mainly dairy cattle and pig improvement programs (Clark, et al., 2012). The aim of this study was to find out the influence of the genomic data on the showjumping horses performance evaluation for the purpose of genetic evaluation and subsequent estimation of breeding values of sport horses in Slovakia.

Material and Methods

For the evaluation of showjumping performance data of Global Champions Tour (season 2012) was used. Competitions were performed in twelve places in Europe and in Middle East. We used data from following levels of competition: 110, 115, 120, 125, 130, 135, 140, 145, 150, 155 and 160 centimeter (CSI1*, CSI2*, CSI5* and Special Invitational). The 4,729 individual competitions were recorded. The age of involved horses (1,021 horses) was from 6 to 20 years. From evaluation we excluded those horses, which were disqualified, had not recorded profit penalties or time from competitions. The simulation of single nucleotide polymorphisms (SNPs) was performed by SNP_SIM program. It is a simple simulator of SNP genotypes and phenotypes developed by Misztal (2009). The fourteen SNPs linked with showjumping performance (Gu, et al., 2009; Hill, et al., 2010; Schröder, 2010; Trakovická, et al., 2012) were used for the simulation analyse. Basic statistical analyses were performed in the program packages SAS (THE SAS SYSTEM V 9. 2.). We used the GLM procedure, the RANK procedure in SAS 9. 2. for further detailed analyses. Input data was not normally distributed. We practised transformation to the normally distributed variable ranking of horses by Blom score (height of obstacles and SNP effect taken into account). Model equations used to the estimation of the relationship between the factors and the ranking of horses:

Model 1 (without SNP effect)

$$Y_{ijklmno} = \mu + B_i + P_j + C_k + A_l + PP_m + T_n + e_{ijklmno}$$

Model 2 (without SNP effect – effects age, penalty points and time in form of quadratic regression)

$$Y_{ijklmno} = \mu + B_i + P_j + C_k + b(age)_{ijkl} + c(points)_{ijklm} + d(time)_{ijklmn} + e_{ijklmno}$$

Model 3 (with SNP effect, effect age, penalty points and time in form of quadratic regression)

$$Y_{ijklmnop} = \mu + B_i + P_j + C_k + SNP_l + b(age)_{ijklm} + c(points)_{ijklmn} + d(time)_{ijklmno} + e_{ijklmnop}$$

where: Y dependent variable ranking of horses; μ mean value; B_i fixed effect of breed; P_j fixed effect place of race; C_k fixed effect of competition; A_l fixed effect of age; PP_m fixed effect of penalty points; T_n fixed effect of time; SNP_l is the effect of the simulated single nucleotide polymorphism (different count of used SNP); b(age) effect age in form of quadratic regression; c(points) effect penalty points in form of quadratic regression; d (time) effect time in form of quadratic regression; e random error related with observation Y.

Results and discussion

Sport horses are evaluated for their performing ability, or more specifically, their ability to win competitions. These traits are partly inherited from their parents and partly due to nutrition (Gálik, et al., 2011; Gálik, et al., 2012), training and other non-genetic factors. In animal breeding, these are referred to as "genetic" and "environmental" factors respectively. To explain the variability of horses ranking we used some different models. In the first model without SNP effect we are able to describe 53 % of performance (coefficient of determination R^2 was 0.53). All used fixed effects were high statistically significant, except the effect competition (Table 1).

Table 1. Linear model without SNP effect
Tabuľka 1. Lineárny model bez SNP efektu

R^2 0.53

Source	**F Value**	**Pr**
Breed	3.73	0.0001***
Place of race	20.61	0.0001***
Competition	4.80	0.0025**
Age	13.06	0.0003***
Penalty Points	3013.97	0.0001***
Time	161.18	0.0001***

*P < 0.05, **P < 0.01, ***P < 0.001

In the case of using of linear model with some effects in form of quadratic regression (age, penalty points and time) we obtained coefficient of determination 0.60. All effects were high statistical significant in this model, except effect age in form of quadratic regression (Table 2).

Table 2. Linear model without SNP effect, some effects in form of quadratic regression
Tabuľka 2. Lineárny model bez SNP efektu, niektoré efekty vo forme kvadratickej regresie

R^2 0.60

Source	**F Value**	**Pr**
Breed	4.38	0.0001***
Place of race	24.18	0.0001***
Competition	5.63	0.0008***
Age	15.33	0.0001***

Age*Age	9.78	0.0018^{**}
Penalty Points	3548.37	0.0001^{***}
Penalty Points*Penalty Points	574.17	0.0001^{***}
Time	115.86	0.0001^{***}
Time*Time	25.87	0.0001^{***}

*P < 0.05, **P < 0.01, ***P < 0.001

Highest coefficient of determination $R^2 = 0.61$ we obtained in the third model where one SNP effect was taken into account. This SNP effect was the most statistical significant to the variable ranking of horses. All used effects were high statistical significant, except effect age in form of quadratic regression and the SNP effect was not statistical significant (Table 3).

Table 3. Linear model with one simulated SNP effect
Tabuľka 3. Lineárny model s jedným simulovaným SNP efektom

R^2 0.61

Source	**F Value**	**Pr**
Breed	4.42	0.0001^{***}
Place of race	24.41	0.0001^{***}
Competition	5.68	0.0007^{***}
SNP	0.23	0.8783
Age	15.39	0.0001^{***}
Age*Age	9.74	0.0018^{**}
Penalty Points	3605.03	0.0001^{***}
Penalty Points*Penalty Points	585.91	0.0001^{***}
Time	117.86	0.0001^{***}
Time*Time	27.22	0.0001^{***}

*P < 0.05, **P < 0.01, ***P < 0.001

Highest coefficient of determination $R^2 = 0.63$ we obtained in the model where we used fourteen simulated SNP effects (Table 4). Effect age of horses in form of quadratic regression and SNP11 effect were not statistical significant. All other effects were high statistical significant or statistical significant.

Table 4. Linear model with fourteen simulated SNP effects, some effects in form of quadratic regression
Tabuľka 4. Lineárny model so štrnástimi simulovanými SNP efektami, niektoré efekty vo forme kvadratickej regresie

R^2 0.63

Source	**F Value**	**Pr**
Breed	4.77	0.0001^{***}
Place of race	25.95	0.0001^{***}
Competition	6.23	0.0003^{***}

SNP1	39.36	0.0001***
SNP2	33.59	0.0001***
SNP3	20.45	0.0001***
SNP4	16.37	0.0001***
SNP5	13.11	0.0001***
SNP6	5.36	0.0011**
SNP7	16.33	0.0001***
SNP8	13.72	0.0001***
SNP9	4.69	0.0029*
SNP10	5.83	0.0006***
SNP11	1.20	0.3078
SNP12	3.07	0.0267*
SNP13	2.76	0.0407*
SNP14	2.87	0.0353*
Age	7.68	0.0056***
Age*Age	3.41	0.0648
Penalty Points	3455.33	0.0001***
Penalty Points*Penalty Points	587.72	0.0001***
Time	100.91	0.0001***
Time*Time	27.92	0.0001***

*P < 0.05, **P < 0.01, ***P < 0.001

It is possible to use simulated SNP data for evaluation of performance of showjumping horses. In our work we used only fourteen simulated SNP effects in the evaluation. We observed increase of coefficient of determination in dependence on number of involved SNPs in the model. In case of fourteen simulated SNP effects used in the linear model, increase of coefficient of determination was 0.03 compared to the model without SNP effect. This value is not high. It is possible to obtain increase of coefficient of determination and more accurately estimating of breeding values involving more SNPs to the evaluation. Using a new SNP chip (Equine SNP50), which includes 54,602 SNP markers distributed among the whole equine genome, other genome-wide scan projects are ongoing to detect SNP markers of showjumping and endurance race ability (Barrey, 2010). The genomic selection based on single nucleotide polymorphism allows us to select juvenile animal in a greater accuracy in comparison with the traditional genetic evaluation systems. It can significantly increase the accuracy of pre-selection of breeding animals (Meuwissen, et al., 2001). Simulated SNP effects were used by Riecka and Candrák (2010) to influence the genomic data in the breeding values estimation in national Holstein cattle population in the Slovak republic. Including of genomic data into the estimation of breeding values could be significant. A significant higher coefficient of determination about 5 % estimated by the linear model was observed when the SNP effects were included. VanRaden (2012) obtained similar results in dairy cattle population. Several of genomic chips with different densities of SNPs were used. Reliability increased using chips with height of densities SNPs. There are many different ways how to use simulated genomic data. Clark, et al. (2012) used simulated and real data to examine the effects of various degrees of relationship on the accuracy of genomic selection. Genomic Best Linear Unbiased Prediction (gBLUP) was compared to two pedigree based BLUP methods, one with a shallow one generation pedigree and the other with a deep ten generation pedigree. The gBLUP method predicted breeding values more accurately than BLUP.

Conclusions

Our results clearly indicated that for the sport horses evaluation it is possible to obtain more accurately estimating of breeding values involving SNPs to the estimation. In our work we used only fourteen simulated SNP effects in the evaluation. Increase of coefficient of determination was 0.03 compared to the model without SNP effect. Using a new SNPs chip (Equine SNP50) it is possible to involve more SNPs into the evaluation and obtained more accurately breeding values.

Acknowledgements

We thank Angelique van Eijk - Social Media Expert at Global Champions Tour for agreement to use the data from Global Champions Tour season 2012.
This work has been supported by the grants:
The Grant Agency of the Ministry of Education Slovak Republic (Grant KEGA No. 027SPU 4/2012; Grant APVV No. 0636-11) and the Excellence Center for Agrobiodiversity Conservation and Benefit project (ITMS: 26220120015) implemented under the Operational Programme Research and Development financed by the European Fund for Regional Development.

References

Barrey, E., (2010) Genetics and genomics in equine exercise physiology: an overview of the new applications of molecular biology as positive and negative markers of performance and health. Equine veterinary Journal, 42 (38), 561-568. DOI: 10.1111/j.2042-3306.2010.00299.x

Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W., Veerkamp, R.F. (2008) Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. Genetics, 178 (1), 553-561. DOI: 10.1534/genetics.107.080838

Clark, S.A., Hickey, J.M., Daetwyler, H.D., van der Werf, J.H.J. (2012) The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genetic Selection Evolution 44 (4), 9. DOI:10.1186/1297-9686-44-4

Daetwyler, H.D., Calus, M.P.L., Pong-Wong, R., de los Campos, G., Hickey, J.M., (2013) Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting and Benchmarking. Genetics, 193 (2), 347-365. DOI: 10.1534/genetics.112.147983

Gálik, B., Bíro, D., Halo, M. Juráček, M., Šimko, M., Massanyi, P., Rolinec, M., (2012) The effect of different macromineral intakes on mineral metabolism of sport horses. Acta Veterinaria Brno, 81 (2), 113-117. DOI:10.2754/avb201281020113

Gálik, B., Bíro, D., Halo, M., Juráček, M., Šimko, M., Vaščáková, V., Rolinec, M., (2011) The effect of phytoadditives on macroelements digestibility of sport horses. Journal of Central European Agriculture, 12 (3), 390-397. DOI:10.5513/JCEA01/12.3.928

Gu, J., Orr, N., Park, S. D., Katz, L. M., Sulimova, G., MacHugh, D. E., Hill, E. W., (2009) A Genome Scan for Positive Selection in Thoroughbred Horses. PloS One 4, 1-17. [Online] Available from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0005767 (accessed 2 June 2009).

Hill, E. W., Gu, J., McGivney, B. A., MacHugh, D. E., (2010) Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. Animal Genetics, 41(2), 56-63. DOI:10.1111/j.1365-2052.2010.02104.x

Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E., (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157 (4), 1819-1829.

Misztal, I., (2009) SNP_Sim ver. 1.2. Simulation and analyzes of genomic SNP data, University of Georgia

Riecka, Z., Candrák, J. (2010) The influence of the genomic data on the breeding values estimation in the Slovak republic. Acta fytotechnica et zootechnica 13 (2), 33-35.

SAS Institute Inc. (2008). SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc.

Schierenbeck, S., Pimentel, E. C. G., Tietze, M., Körte, J., Reents, R., Reinhardt, F., Simianer, H., König, S., (2011) Controlling inbreeding and maximizing genetic gain using semi-definite programming with pedigree-based and genomic relationships, Journal of Dairy Science, 94 (12), 6143-6152. DOI: 10.3168/jds.2011-4574

Schröder, W., (2010) Athletic performance and conformation in Hanoverian warmblood horses – population genetic and genome-wide association analyses. Dr. Thesis. University of Veterinary Medicine Hannover, Hannover.

Sonesson, A.K., Woolliams, J.A., Meuwissen, T.H.E., (2010) Maximising Genetic Gain Whilst Controlling Rates Of Genomic Inbreeding Using Genomic Optimum Contribution Selection. Proceedings of 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany, August 1-6.

Trakovická, A., Gábor, M., Miluchová, M., Minarovič, T., Šťastná, D., (2012) Analysis of the Nebulin-Related Anchoring Protein Gene (N-RAP) SNP Polymorphism (C/T) in Slovak Warmblood Horse by PCR-RFLP Method. Animal Sciences and Biotechnologies, 45 (1), 265-268.

VanRaden, P. M., Olson, K. M., Null, D. J., Sargolzaei, M., Winters, M., van Kaam J. B. C. H. M., (2012) Reliability Increases from Combining 50,000- and 777,000-Marker Genotypes from Four Countries. Proceedings of Interbull Annual Meeting. Cork, Ireland, May 28-June 1.