56th Annual Meeting of the EAAP, Uppsala, Sweden, June 5th - 8th , 2005.
Horse Commission. Session1a. Systems of Identification in Horses.

Connectedness among five European sport horse populations - aspects on ID recording and exchange of pedigree data

E. Thorén, H. Jorjani and J. Philipsson, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, S-750 07 Uppsala, Sweden.*

Abstract

While investigating the genetic ties between five European studbooks in order to check the feasibility of international genetic evaluation of sport horses, difficulties related to ID-numbers of stallions were encountered. It turned out that the original ID-number of a stallion was very seldom registered when he was used in breeding in another studbook area than his studbook of birth. To determine the extent of genetic ties among the horse populations the concept of genetic similarity (GS) was employed. This measures the number of tested horses sired by stallions with progeny in two organisations in relation to the total number of tested animals in those two organisations. Data on sport horse populations were provided from Denmark (DWB), Hanover (HAN), Holstein (HOLST) the Netherlands (KWPN) and Sweden (SWA), including 2381 stallions which together had 64225 progenies in these five populations. The GS values varied between 7 % (NLD and SWE) and 29 % (HAN and NLD). The average GS value among the horse populations was 15 %. This is almost twice as high as has previously been reported for major dairy cattle breeds. Considering these results international genetic evaluations of sport horses seem feasible. However, a major obstacle is the lack of unique ID numbers of horses across countries.

Introduction

Correct identity and pedigree recording is fundamental for all animal breeding activities. It becomes especially important when methods used for genetic evaluation take the performance records of relatives into account. Today's animal breeding is not limited by geographical areas and scarcely by time, which have stressed the importance of unique and universal identities of animals across breeds and countries. While working with computerised pedigree data it is essential to know that the information is correct, unique and compatible between databases. Errors in the pedigree can have severe effects on the accuracy of the genetic evaluations and eventually result in a slower genetic progress than would have been obtained with correct pedigrees. Furthermore, misidentification of horses may have severe economic and legal consequences. Effects of erroneous pedigrees have been shown for example in cattle (Visscher et al., 2002; Banos et al., 2001; Van Vleck, 1970a and 1970b) and swine (Long et al., 1990). In commercial dairy cattle breeding, paternity misidentification rates as high as 10% have been verified (Banos et al., 2001). There are various sources of pedigree errors. Potential practical sources are mix-up of transported semen from different stallions and confusion of foals with their dams in flocks of broodmares. On the administrative side confusion may occur when merging datasets from different origins and ID-records are not unique or compatible between computer systems. Such merging takes place when the necessary information on animals for genetic evaluations or other purposes is not initially reported to the same organisation.

With the improved techniques for the storage and transport of semen during the latest decade, the use of certain stallions in several countries/organisations has increased markedly. The increased use of foreign stallions has created a demand for objective information on stallions in many countries, both among breeders and breed organisations. This demand has in turn lead to the formation of the international working group Interstallion, under the aegis of the World Breeding Federation for Sport Horses (WBFSH). One of the main objectives for Interstallion is to investigate the possibilities of international genetic evaluation of stallions.

If data are going to be used from several sources (data bases of different breed societies) the first and indispensable prerequisite to perform an international genetic evaluation is that genetic connectedness, i.e. the genetic ties, among participating populations are at an acceptable level. Otherwise it is not possible to estimate the genetic correlation between results obtained in any combination of two breeding organisations. Such correlations are necessary to assess the similarity in traits being evaluated in different countries or breeding organisations. Hanocq et al. (1996) showed that lack of connection may introduce a large bias in the estimation of genetic level of subpopulations, because the differences among subpopulations are not estimable and are assumed to be zero. Differences in genetic level become possible to estimate even with a rather limited degree of genetic connectedness, but the accuracy of the estimation increases with the degree of connectedness. Also for the genetic evaluation of individuals the accuracy increases with increasing connectedness. Hence, with low connectedness, which still is larger than zero, international genetic evaluation may be feasible, but not recommended, because there will be too much fluctuations in the results.

To determine the extent of genetic ties among the horse populations the concept of genetic similarity (Rekaya et al. 2003) was employed. Genetic similarity has been previously used as an indication of balancedness of data and hence connectedness in dairy cattle populations (Jorjani, 2000). For this purpose data from five European horse populations were used. These five organisations have been shown to have similar types of young horse performance data and the hypothesis was that they have an overlapping use of stallions. To find common stallions, i.e. stallions with offspring in more than one population, stallion names and birth year together with the data on their sires and maternal grandsires were used.

Materials and methods

Five sport horse breeding organisations were asked to provide information on stallions with tested progeny, preferably from tests included in the national genetic evaluations. The request comprised pedigree information and number of tested offspring per stallion. A detailed list of the requested data is provided in appendix 1. Tested progeny was set as a prerequisite in order to ensure that the stallions contributed actively with genetic material to the populations. The organisations contributing with data for the study were; the Danish Warmblood Society (DWB), the Hannoveraner Verband from Germany, the Holsteiner Verband from Germany, the Royal Warmblood Studbook of the Netherlands (KWPN) and the Swedish Warmblood Association (SWA).

Material

The number of stallions with tested progeny in the original data sent from the organisations as well as in the edited data used for estimations of genetic connectedness is shown in table 1. The number of tested progeny of each organisation in the original and the edited data are also shown in table 1. Table 2 shows the source of data and the time period during which the progenies were tested.

Table 1. Number of stallions with tested progeny and the number of progeny in the original data sent from the organisations and in the edited data

Organisation	**DWB**	**Hanover**	**Holstein**	**KWPN**	**SWA**
No of stallions in original data	807	762	300	452	750
No of stallions in edited data	609	762	300	453	749
No of progeny in original data	24938	14950	4987	2993	16371
No of progeny in edited data	24924	14950	4987	2993	16371

Table 2. Source of data and the time period during which the progeny were tested

Organisation	**Source of data**	**Time period**
DWB	Various tests[1]	-2003
Hannover	"Zucht-stuten prüfung"= field test for young mares	1986 - 2002
Holstein	"Zucht-stuten prüfung"= field test for young mares	1986 - 2002
KWPN	stallion performance test and 5-week mare test at station	1980 - 2003
SWA	RHQT[2]	1973 - 2003

[1] Foal inspections, young horse tests, stallion performance tests, etc.
[2] Riding Horse Quality Test = field test for young horses, all genders

Methods

Initially, the idea was to use original ID-numbers to retrieve common stallions between organisations. Very soon it turned out that this would result in severe underestimation of the genetic connectedness, because the original ID:s of imported stallions (and their pedigree information) had been altered to new ID:s in almost all cases. Hence, a decision was taken to use the name of stallions instead. However, name is also a somewhat unreliable parameter to work with, since it is common with several possible spellings of the same name, horses change names due to changes of owner or sponsor and different stallions with the same name occur in several organisations. Additionally, sometimes full brothers are named identically and only separated by a number in connection to the name. As an example of stallions having different names and numbers in different organisations, the famous Holsteiner stallion Ramiro has been given the names and numbers listed in table 3 in the provided material. Another example is the French-born legendary stallion Cor de la Bryère (table 3). Note that his French original ID isn't used in any of the organisations.

Table 3. Different names and numbers on two well-known stallions

Organisation	Names	ID-numbers	Names	ID-numbers
Holst	RAMIRO	321210389565	Cor d.l.Bryere	321210398168
KWPN	G. RAMIRO	K294STB-H	Cor de la Bryere	K21 0398168
SWA	RAMIRO Z	-2714	Cor de la Bryère	-8060

In our continued editing procedure name of stallion was considered the best alternative in combination with sire, maternal grandsire and birth year, when ID-numbers could not be used. Extensive editing of the material followed and some editing information is listed in appendix 2.

Before editing, the material included 3017 individual stallions. After editing, this number was reduced to 2381 stallions. The reduced number is explained by the fact that one individual could show up as several individuals before editing, because of differences in spelling etc. in different organisations. The edited material was used for comparison, pair wise between organisations. Matching was done in four steps:

Match on name (or part of name)
Match on name + sire's name
Match on name + sire's name + maternal grandsire's name
Match on name + sire's name + maternal grandsire's name + Birth year of stallion

While editing and matching, a number of oddities in the provided material were found. For example several stallions had no birth year registered, or different birth year in different organisations, even if it obviously was the same individual. Such oddities can lead to the classification of the same stallion as two different individuals in the merging procedure and consequently loosing a link between the two populations. Most often it requires visual inspection to detect these flaws and to correct them, which is a very time-consuming process.

The Genetic Similarity (GS) was calculated according to the formula introduced by Rekaya et al. 2003. Compared to the number of common stallions between organisations the GS is a more relative measure (expressed in percents) that takes into account the number of tested progeny in each organisation.

Results

Using only number of common stallions (table 4) or common tested progeny (table 5) as a measure of the genetic ties between breed organisations indicated that the strongest ties should be between SWA and DWB followed by Holstein and DWB. However, combining these numbers into the GS (table 6) showed stronger genetic ties between Hanover and KWPN than the others. Lowest GS value was found between KWPN and SWA. The mean GS value for the participating organisations were: Hanover 16.7 %, Holstein 13.6 %, KWPN 14.0 %, DWB 16.8 % and SWA 11.7 %. This indicates that DWB has the strongest genetic ties to the other breed organisations, while SWA has the weakest. It is then important to keep in mind that the time period of tested progeny influences the GS values. No restriction in time period of tested progeny or birth year of participating stallions were imposed in this study. Because the progeny test period of the SWA began 7 years before the progeny test period of KWPN and 13 years before the two German organisations, it's reasonable to believe that this influences the mean GS value for SWA in a negative way. Unlike the data from the other organisations, the data from DWB includes progeny tested in foal tests. Consequently stallions approved for breeding as late as 2002 might be present in the Danish data. With the current fast increase in foreign stallion exchange between organisations, the presence of foal test results in the Danish material is likely to influence the mean GS value for DWB in a positive way.

Table 4. Common stallions. On the diagonal, number of stallions with tested progeny in each org. Above the diagonal number of common stallions pair wise between organisations

	DWB	**Hanover**	**Holstein**	**KWPN**	**SWA**
DWB	609	69	103	84	117
Hannover		300	57	44	39
Holstein			762	32	54
KWPN				453	63
SWA					749

Table 5. Common offspring. On the diagonal, number of tested progeny in each org. Above the diagonal, number of offspring to sires having offspring in both organisations

	DWB	**Hanover**	**Holstein**	**KWPN**	**SWA**
DWB	24924	4654	7706	2412	9750
Hannover		4987	2996	2316	1571
Holstein			14950	2040	2720
KWPN				3016	1328
SWA					16371

Table 6. Genetic similarity among organisations in %

	DWB	**Hanover**	**Holstein**	**KWPN**	**SWA**
DWB	100	15.6	19.3	8.6	23.6
Hannover		100	15.0	28.9	7.4
Holstein			100	11.4	8.7
KWPN				100	6.9
SWA					100

Discussion

The average GS value among these horse populations was 15.0 %. This is almost twice as high as has previously been reported for major dairy cattle breeds (Jorjani, 2000). Considering these results a computerized international genetic evaluation of sport horses seems feasible. However, a major obstacle for such a development is the lack of unique ID numbers of horses across countries.

In this study, we have tried to match stallions on names, which is a great risk with all different spellings, changes of names, sponsor's names etc. To use names to retrieve common stallions is not a sustainable alternative in the future work with an international genetic evaluation because of this risk, not to mention the time-consuming process of editing the data. Exchanging information on individuals between organisations would become easier with some kind of uniform identification system. In the future the Universal Equine Life Number (UELN), brought into use about three years ago (WBFSH, 2004), seems like the solution to many problems. Still, most organisations are implementing the UELN successively, which means that we will have to deal with the ID-problems for a long time yet. On the other hand, if the UELN is to be given also to older animals as well as pedigree animals no longer alive, this has to be done with utmost care. Otherwise there is an obvious risk that pedigree errors become permanent and some individuals become registered more than once. The most important measure in any case would be to always keep and use the original ID of a horse! A stallion could of course be given additional studbook numbers in every organisation he is used, which might facilitate the inter-organisational work. At any data exchange however, his original ID should always be used!

During the work with this part of the Interstallion Pilot Project I, a number of important experiences have been made; As the initial intention was to use ID numbers to retrieve common animals among organisations, the request of material did not include names of pedigree animals, which will affect the subsequent studies of genetic connectedness. Another detail missing in the request was test year of progeny or time span of the testing period, which also influence the degree of connectedness. To avoid similar problems in the later pilot studies, as well as in future routine international genetic evaluations of stallions, an international standard for reporting of material should be designed.

References

Banos, G., Wiggans, G.R., Powell, R.L. 2001. Impact of Paternity errors in cow identification on genetic evaluations and international comparisons. *J. Dairy Sci.* 84:2523-2529.

Hanocq, E., Boichard, D., Foulley, J.L. 1996. A simulation study of the genetic effect of connectedness on genetic trend. Genet. Sel. Evol. 28:67-82.

Jorjani, H. 2000. Well-connected, informative sub-sets of data. *Interbull bullentin* 25:22-25.

Long, T.E., Johnson, R.K., Keele, J.W. 1990. Effects of errors in pedigree on three methods of estimating breeding value for litter size, backfat and average daily gain in swine. *J. Anim. Sci.* 68:4069-4078.

Rekaya, R., Weigel, K.A., Gianola, D. 2003. Bayesian estimation of a structural model for genetic covariances between milk yield in five regions of the USA.. *J. Dairy Sci.* 86:1837–1844.

Van Vleck, L. D. 1970a. Misidentification in estimating the paternal-sib correlation. *J. Dairy Sci.* 53:1469–1474.

Van Vleck, L. D. 1970b. Misidentification and sire evaluation. *J. Dairy Sci.* 53:1697–1702.

Visscher, P.M., Woolliams, J.A., Smith, D., Williams, J.L. 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *J. Dairy Sci.* 85:2368-2375.

Internet reference

WBFSH. 2004. www.wbfsh.org. Accessed 2005-05-19

Appendix 1. List of requested material

Information was requested on all stallions which had been in use during the last period of time, covering at least two full generations.

ID number stallion
name (capitals)
id number sire
id number dam
id number of maternal grandsire
id number of paternal grandsire
date of birth stallion (DDMMYYYY)
Unique Equine Life Numbers (UELN), if existing
ID number of the sire in the country of birth

Appendix 2. Editing information

All letters transformed into capitals.

All additional information that sometimes was included in the space reserved for "stallion name", such as id-number, earlier names, sponsor names etc, were removed from the name, and if appropriate, moved to another space.

Where there were several full brothers listed with roman numbers (I, II, III etc.) in connection with the name, the first brother (number I) was named without the roman number I.

Symbols for thoroughbred (xx), Anglo Arab (x or AA) and Arab (ox) in connection with the name or any other symbols or abbreviations of which the only purpose was to indicate breed were removed.

Spelling

Original letter	Replaced with
É	E
È	E
Ü	U
Å	AA
Æ	AE
Ä	AE
Ø	OE
Ö	OE