Microsatellite allele imputation from SNP genotypes for parentage verification in sport horses

H Crichan, C Engler, E Goulas, S Dhorne-Pollet, M Addes, Anne Ricard

▶ **To cite this version:**

H Crichan, C Engler, E Goulas, S Dhorne-Pollet, M Addes, et al.. Microsatellite allele imputation from SNP genotypes for parentage verification in sport horses. World Congress on Genetics Applied to Livestock Production, Jul 2022, Rotterdam (NL), Netherlands. ⟨hal-04650364⟩

HAL Id: hal-04650364
https://hal.inrae.fr/hal-04650364v1

Submitted on 16 Jul 2024

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.

HAL Authorization

Microsatellite allele imputation from SNP genotypes for parentage verification in sport horses.

H. Crichan[1*], C. Engler[1], E. Goulas[1], S. Dhorne-Pollet[2], M. Addes[1], A. Ricard[1, 2]

[1]Institut Français du Cheval et de l'Equitation, Pôle développement, innovation et recherche, La jumenterie, 61310 Exmes, France, [2]GABI, INRAE, AgroParisTech, Université Paris Saclay, 78350 Jouy en Josas, France; *harmony.crichan@ifce.fr

Abstract

In the major livestock species, parentage verification is done using Single Nucleotide Polymorphism (SNP) markers. ISAG is working to produce a SNP panel for parental verification in horses. There will be a transition period between microsatellites (MS) and SNPs. Without any tool for the transition, breeders will have to double genotype foals or re-genotype mares and stallions with SNPs. We wanted to produce a tool to predict MS genotypes using SNP markers that could be included on a chip with the new ISAG panel. To achieve this we built a correspondence table between MS alleles and SNP haplotypes. We validated it using kfold cross-validation. With the highest SNP density, we obtained 3.4% of incompatibilities over 2 or more MS or 96.6% of parental verification achieved over the major breeds in our study, Selle-Français (SF, 54%), French Trotters (TF, 22%), Arabs (AR, 17%), Anglo-Arabs (AA, 3%) and other saddlebreds (SE, 4%).

Introduction

Nowadays, parental verification in horses is done using MS genotypes. ISAG is currently working on a new SNP panel for parentage testing that will soon replace the microsatellite panel. Lab technologies to produce microsatellite and SNP genotypes are different so genotypes of these markers cannot be obtained at once. The objective of our study is to predict MS genotypes from SNP haplotypes using commercial SNP chips to perform parental verification.

Materials & Methods

Data. IFCE (Institut Français du Cheval et de l'Equitation), INRAe and Fonds Eperon funded research programs such as Genequin, JumpSNP, SelGenEqui, GenEndurance and SoGen that provided horses with SNP genotypes used in this study. MS genotypes of the SNP genotyped horses and their parents from routine parental verification came from SIRE (equine information database). Horses with incompatibilities with their dam or sire or occurring as a parent of an incompatible foal were removed from the reference population. There were 5,892 horses with a genotype over at least one known MS and with SNP genotypes. The breeds were mainly Selle-Français (SF, 54%), French Trotters (TF, 22%), and Arabs (AR, 17%), with few Anglo-Arabs (AA, 3%) and other saddlebreds (SE, 4%).
As the data came from different research programs, horses were genotyped with three different SNP chips: 54K (54,602 SNPs, 29%), 65K (65,157 SNP, 15%) and 670K (670,806 SNP, 56%). SF were mostly genotyped with the 670K SNP chip, half of the TF were genotyped with the 54K and the other half with the 670K, and AR mostly with the 65K.
By investigating the MS genotypes, we observed that three MS over the 14 were missing for most of the horses, we had genotypes for only 11.4% of the horses for ASB17, 11.4% for ASB23 and 10% for HSM2. Result were therefore obtained for the 11 MS remaining

Method – Correspondence table building. The method was based on McClure *et al.* (2012) and McClure *et al.* (2013) who worked on cattle. SNPs located within 500,000 base pairs on each side of the MS were selected. We then phased MS alleles together with SNP genotypes using Fimpute3 (Sargolzaei et al., 2014) by converting MS alleles into pseudo-SNP (Karimi *et al*. 2018, Teissier *et al*. 2020). This transformation was useful because FImpute3 only works bi-allelic markers. From the imputation, we obtained SNP haplotypes and MS alleles alongside. A correspondence table was then created between MS alleles and SNP haplotypes. The table was built so that one MS allele could have multiple correspondences with SNP haplotypes but a SNP haplotype had to correspond to only one MS allele. A second rule was that only the closest SNP constituted the haplotypes. Eight to 40 SNP located on either side of the MS constituted the haplotypes.

Method – Validation. Horses were randomly split in five groups of equal sizes; each group was in turn the validation set and the training set. Using the training set, we built the correspondence table. The MS genotypes of the validation set were predicted using the correspondence table only. The partitioning of the validation sample was repeated 10 times. We then compared the predicted MS alleles with the actual MS alleles information. We measured ineffectiveness using the percentage of MS genotypes attributed with the correspondence table that were different from the original MS genotypes. The second measure of ineffectiveness was the number of MS found incompatible using MS parental information.

Results

Genotypes description. Over the 5,892 horses, parental verification was achievable for 96.1% of them (5,661). Of the 5.9% of the 5,661 horses with at least one MS incompatible, 5.3% had one and 0.6% had two MS incompatible. MS had a variable number of alleles in our dataset, ranging from six to 12, and recruited SNPs around them varied from 111 to 314. 1,908 SNPs were used in this study, the vast majority were in the 670K SNP chip (1,902), and only 10% of them were in the 65K SNP chip.

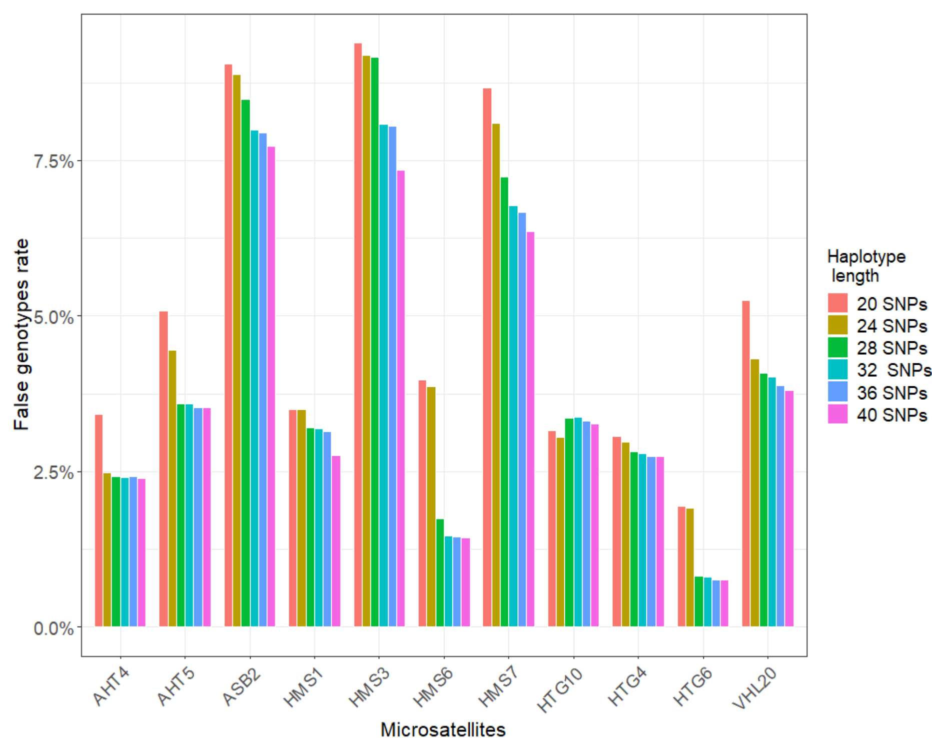

Figure 1: Average rates of false genotypes by MS and by haplotype length in the validation set (10-fold cross validation)

Training and validation sets. Rates of false genotypes were 0.6% for MS HTG6 and 8.1% for MS HMS3 with a 32 SNPs long haplotype. With the k-fold validation, false genotypes varied with the MS considered and with the length of the haplotype with a threshold observed at 32 SNPs (figure 1). The proportion of horses with more than one MS incompatible increased from 0.6% initially to 12.4% with a 32 SNP haplotype, the standard deviation of theses proportions between the random draws was 0.3%. For a 32 SNPs haplotype, we obtained 25% of one MS incompatible and 12% of 2 MS incompatible.

The error rate depended on the breed. Horses with more than one MS incompatible varied from 27.3% for AR to 7.3% for SF. A major difference between these breeds was the chip mainly used for the SNP genotyping.

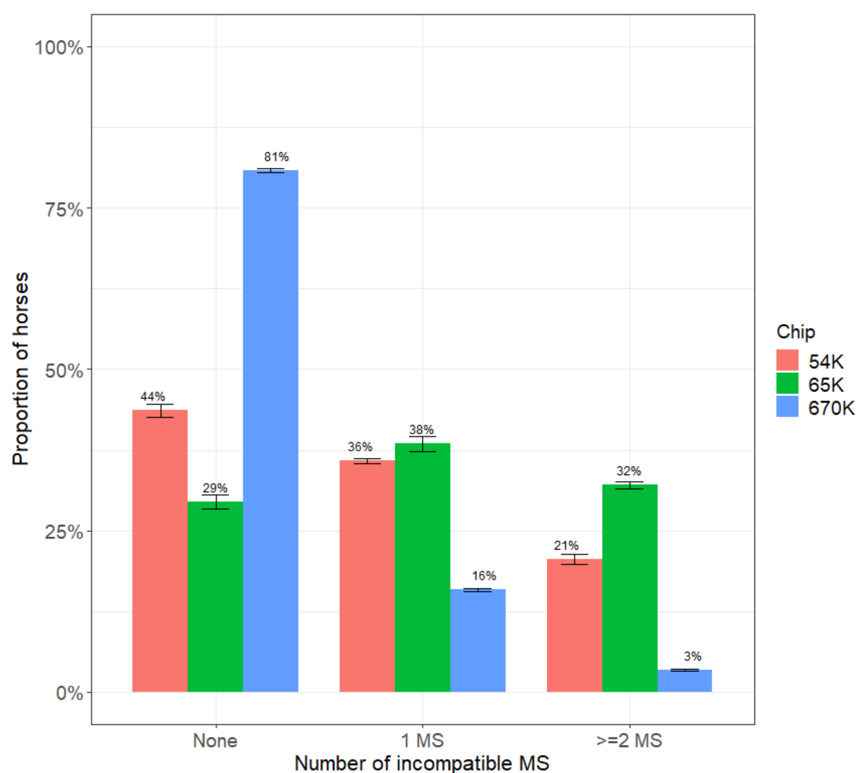

Figure 2: Average rates of incompatible MS by SNP chips in the validation set (10-fold cross-validation)

As figure 2 shows, the incompatibility differences were related to the density of the SNP chip used to genotype the horses. Only 3.4% of horses genotyped with the 670K SNP chip had more than one MS incompatible, whereas this percentage greatly increased to 32.1% and 21% with horses genotyped with the 65K SNP chip and the 54K SNP chip respectively. For horses genotyped with the 670K SNP chip, results were promising on the three breeds with only 4% for FR, 3.4% for AR and 3.2% for SF of horses with more than one MS incompatible.

Discussion

The method of SNP based MS imputation is recommended as a transition tool to save genotyping extra costs. There are some limitations regarding the results obtained from the correspondence table. First, there were three breeds in our study and we saw that the distribution of MS alleles differed across breeds; if we want to be able to predict MS alleles over all breeds, we will need more genotypes from different breeds to improve the corresponding table. This procedure may be difficult to manage and to automate, as some horses will need re-genotyping with MS. It will need to be accepted by breeders who are the primary clients for parental verification. If not, parental verification will need to be done by re-genotyping and/or double genotyping during the transition period. One of the questions is also the value of these results internationally. Horses that produce offspring or perform internationally will most likely be genotyped with both types of marker sets. Moreover, ISAG recommends only the re-genotyping in the first place as it was eventually done with cattle during their transition from MS to SNP. Finally, breeders' organizations will have to encourage their members to SNP genotype their horses so that the transition period is reduced to a minimum.

References

Karimi Z, Sargolzaei M, Robinson J.A.B, and Schenkel F.S. - (2018). Can. J. Anim. Sci. 4:98. 750-759 https://doi.org/10.1139/cjas-2018-0009

McClure M, Sonstegard T, Wiggans G, Van Tassell CP. - (2012) Front Genet. 3:140. https://doi:10.3389/fgene.2012.00140

McClure MC, Sonstegard TS, Wiggans GR, et al. - (2013) Front Genet. 4:176. https://doi:10.3389/fgene.2013.00176

Sargolzaei M, Chesnais J.P, Schenkel F.S. - (2014) BMC Genom. 15:478. https://doi.org/10.1186/1471-2164-15-478

Teissier M, Larroque H., Brito L.F., Rupp R., Schenkel F.S. et al - (2020) JDS. 103 :12 :11559-11573 https://doi.org/10.3168/jds.2020-18662