RÉPUBLIQUE
FRANÇAISE
*Liberté
Égalité
Fraternité*

institut français
du **cheval**
et de l'**équitation**

MICROSATELLITE ALLELES IMPUTATION FROM SNP GENOTYPES FOR PARENTAL VERIFICATION IN SPORT HORSES

CRICHAN H.[1], ENGLER C.[1], GOULAS E.[1], POLLET S.[2], ADDES M.[1], RICARD A[1,2].
[1]: INSTITUT FRANÇAIS DU CHEVAL ET DE L'ÉQUITATION (IFCE)
[2]: INSTITUT NATIONAL RECHERCHE POUR L'AGRONOMIQUE, L'ALIMENTATION ET L'ENVIRONNEMENT (INRAE)

06/12/2021

Context and objectives

Context:

- **SNP markers are more accurate than microsatellite markers (MS) for parentage verification and offer more : major genes, genomic evaluation…**
- **ISAG is currently working on a SNP panel for parentage verification, replacing the MS panel**
- **Currently, mares and stallions are genotyped with MS. New genotyping of foals will be with SNPs. Without any tool, there will be an extra cost in either re-genotyping (SNPs) mares and stallions or in the double genotyping of foals (SNPs+MS)**

Objective:

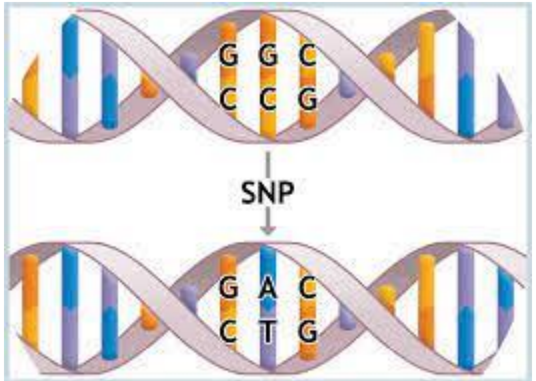
- **Prediction of MS genotypes with SNP haplotypes by building a correspondence table between SNP and MS in order to perform parentage verification with SNP for foals and MS for parents**

Definitions – Microsatellites and SNP markers

Microsatellite marker (MS) or Short Tandem Repeats (STR), here $(CTT)_8$

```
...GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCGCACTTTAACGATACGG...
...CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGAAGCGTGAAATTGCTATGCC...
```

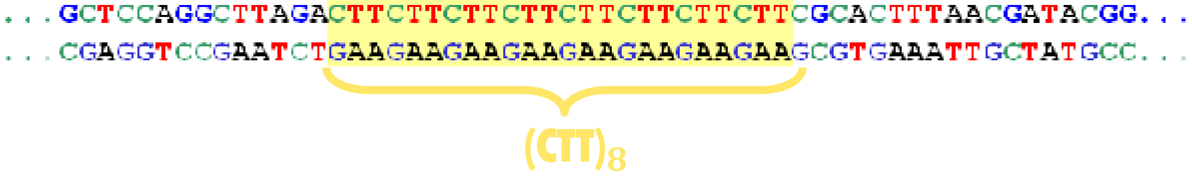
Single Nucleoid Polymorphism (SNP) : 1 locus marker



Method - Microsatellites and SNP locations

- **Microsatellite allele nomenclature defines the number of pattern repetition (ex: $(CTT)_8$)**

Allele K – 8 repetitions



...GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCGCACTTTAACGATACGG...
...CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGAAGCGTGAAATTGCTATGCC...

$(CTT)_8$

Allele J – 7 repetitions

...GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCGCACTTTAACGATACGG...
...CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGCGTGAAATTGCTATGCC...

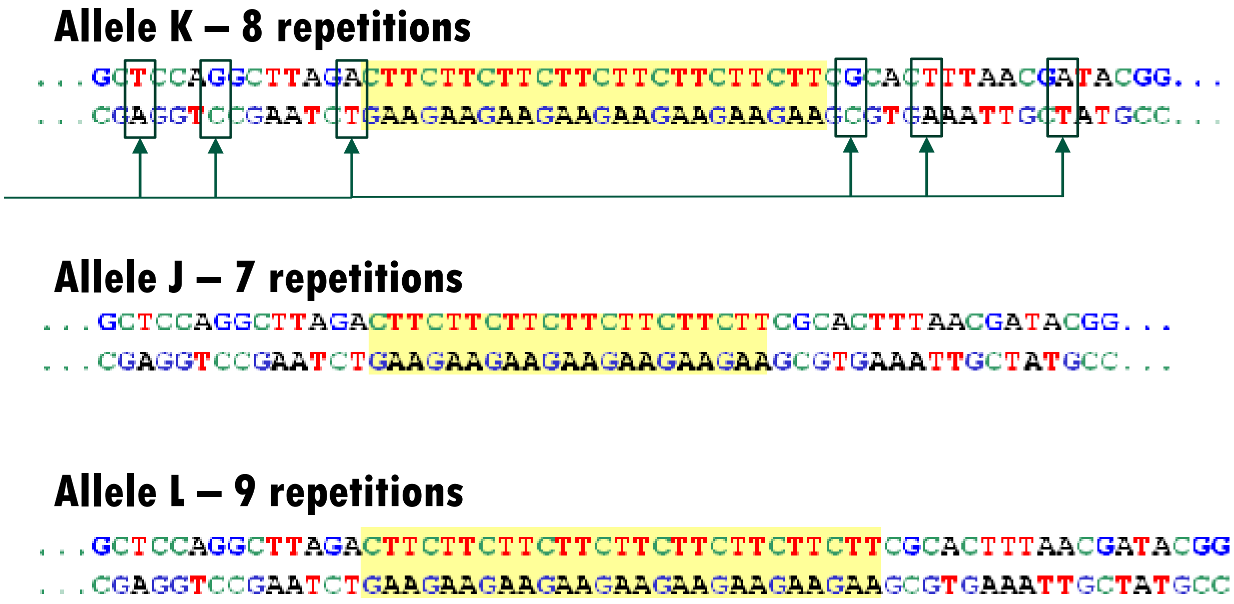
$(CTT)_7$

Allele L – 9 repetitions

...GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCTTCGCACTTTAACGATACGG
...CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGAAGAAGCGTGAAATTGCTATGCC

$(CTT)_9$

Method - Microsatellites and SNP locations



Allele K – 8 repetitions

...GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCGCACTTTAACGATACGG...
...CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGAAGCGTGAAATTGCTATGCC...

Identification of a sequence of SNP close to a MS allele
→ corresponding table

Allele J – 7 repetitions

...GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCGCACTTTAACGATACGG...
...CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGCGTGAAATTGCTATGCC...

SNP were located on either side of each MS +/-500kbp

Allele L – 9 repetitions

...GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCTTCGCACTTTAACGATACGG
...CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGAAGAAGCGTGAAATTGCTATGCC

Method - Microsatellites and SNP locations

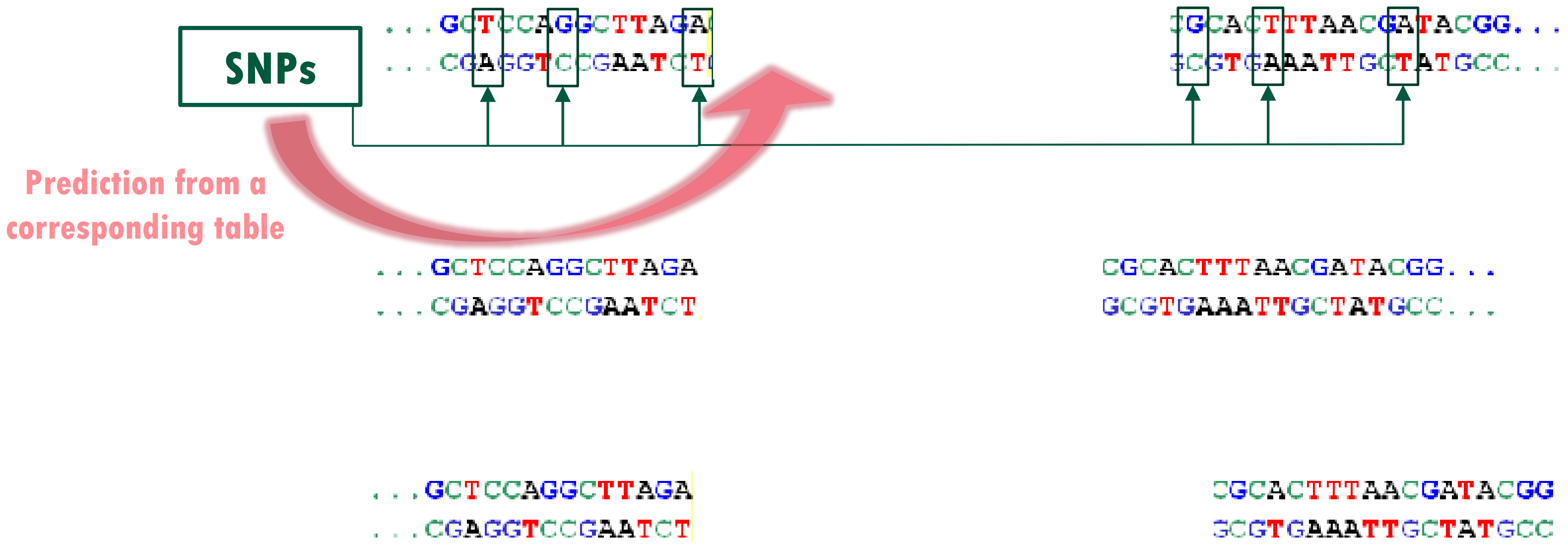
...GCTCCAGGCTTAGA
...CGAGGTCCGAATCT

CGCACTTTAACGATACGG...
GCGTGAAATTGCTATGCC...

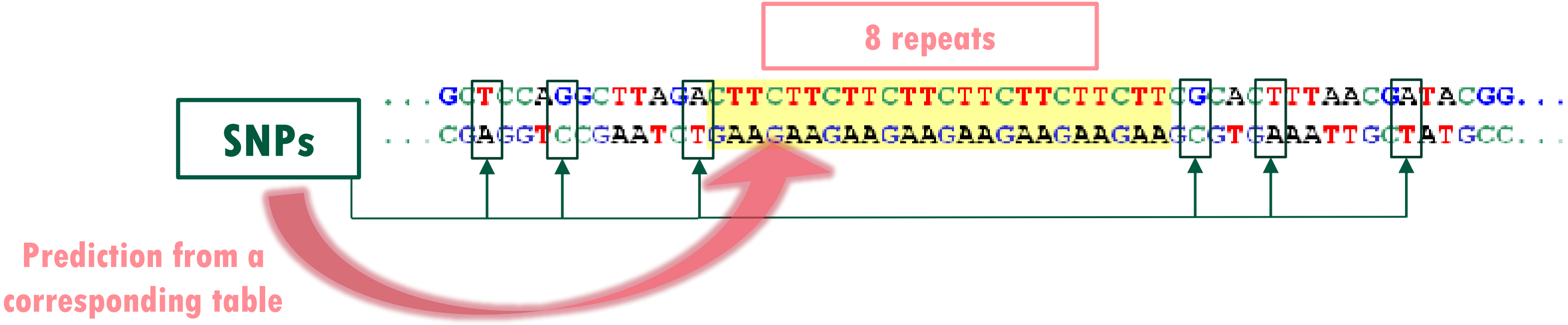
...GCTCCAGGCTTAGA
...CGAGGTCCGAATCT

CGCACTTTAACGATACGG
GCGTGAAATTGCTATGCC

Method - Microsatellites and SNP locations



Method - Microsatellites and SNP locations



8 repeats

...GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCGCACTTTAACGATACGG...
...CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGAAGCGTGAAATTGCTATGCC...

SNPs

Prediction from a corresponding table

...GCTCCAGGCTTAGA
...CGAGGTCCGAATCT

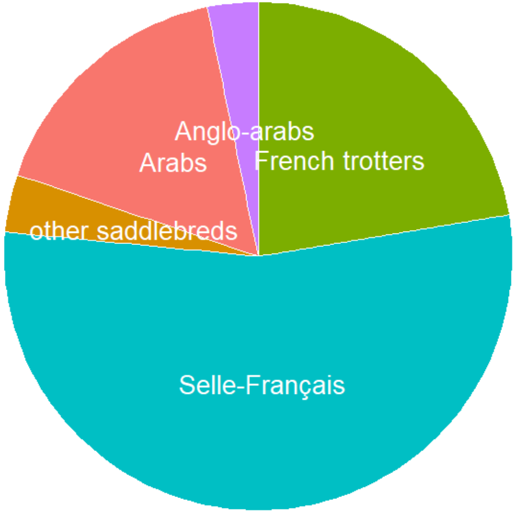
CGCACTTTAACGATACGG...
GCGTGAAATTGCTATGCC...

...GCTCCAGGCTTAGA
...CGAGGTCCGAATCT

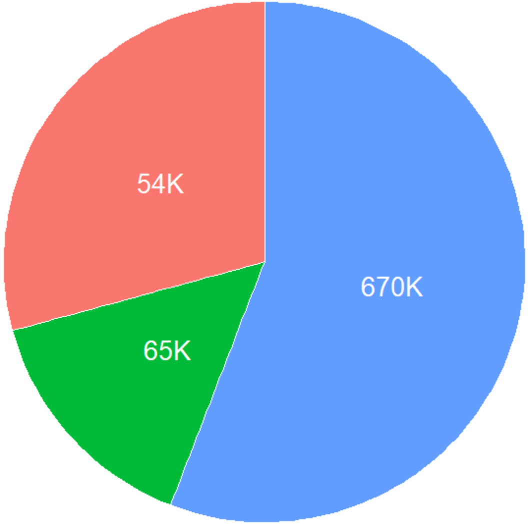
CGCACTTTAACGATACGG
GCGTGAAATTGCTATGCC

Data

- Horses with both SNP genotypes from research programs and microsatellite (MS) genotypes from parental verification
- 5,892 horses (from 5,374 to 5,869 depending on the MS)
- 54 000, 65 000 and 670 000 SNP chips (Illumina/Thermofisher)
- 11 MS (usual reference for parental verification)



Distribution by horse breeds



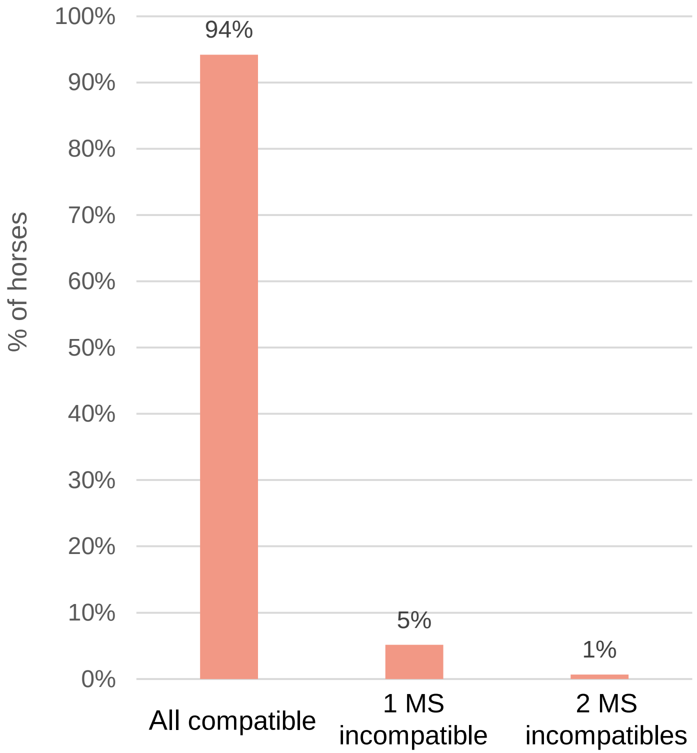
Distribution by size of SNP chip

Validation

- Learning set (80% of dataset) randomly selected from horses with MS + SNP genotypes to create the corresponding table between MS and SNP
- Validation set (20% of dataset) where the MS information is hidden
- We use the corresponding table from the learning set to predict the validation set MS alleles
- Only on the validation set, we count the number of incompatibilities over the 11 MS
- 50 repetitions

Dataset

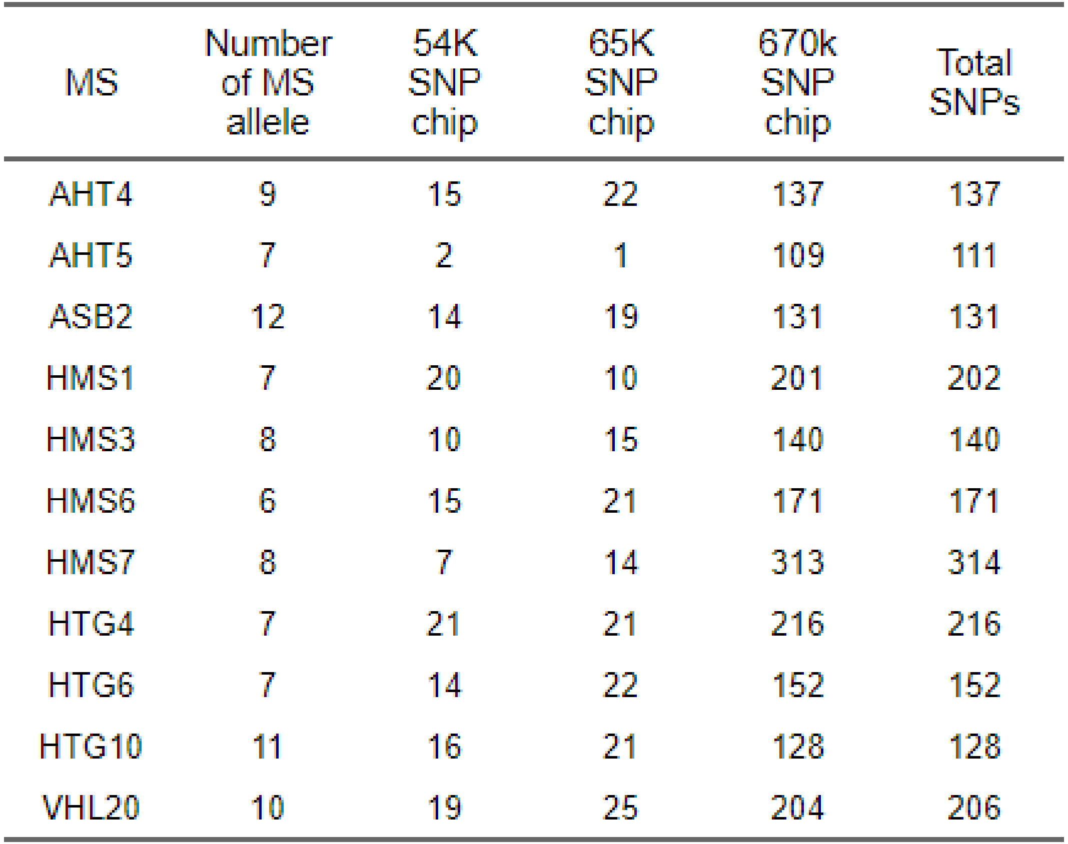
Initially, in the dataset, there were 5% of horses that had a MS genotype incompatible with its parents, 1% of them had 2 MS incompatible (→ rejection of parenthood)



RESULTS : SNPs and MS

MS have a variable number of alleles ranging from 6 to 12, and the number of SNP around them also varies from 111 to 314

A total 1908 SNP were used in this study, the vast majority came from the 670K chip (1902)



MS	Number of MS allele	54K SNP chip	65K SNP chip	670k SNP chip	Total SNPs
AHT4	9	15	22	137	137
AHT5	7	2	1	109	111
ASB2	12	14	19	131	131
HMS1	7	20	10	201	202
HMS3	8	10	15	140	140
HMS6	6	15	21	171	171
HMS7	8	7	14	313	314
HTG4	7	21	21	216	216
HTG6	7	14	22	152	152
HTG10	11	16	21	128	128
VHL20	10	19	25	204	206

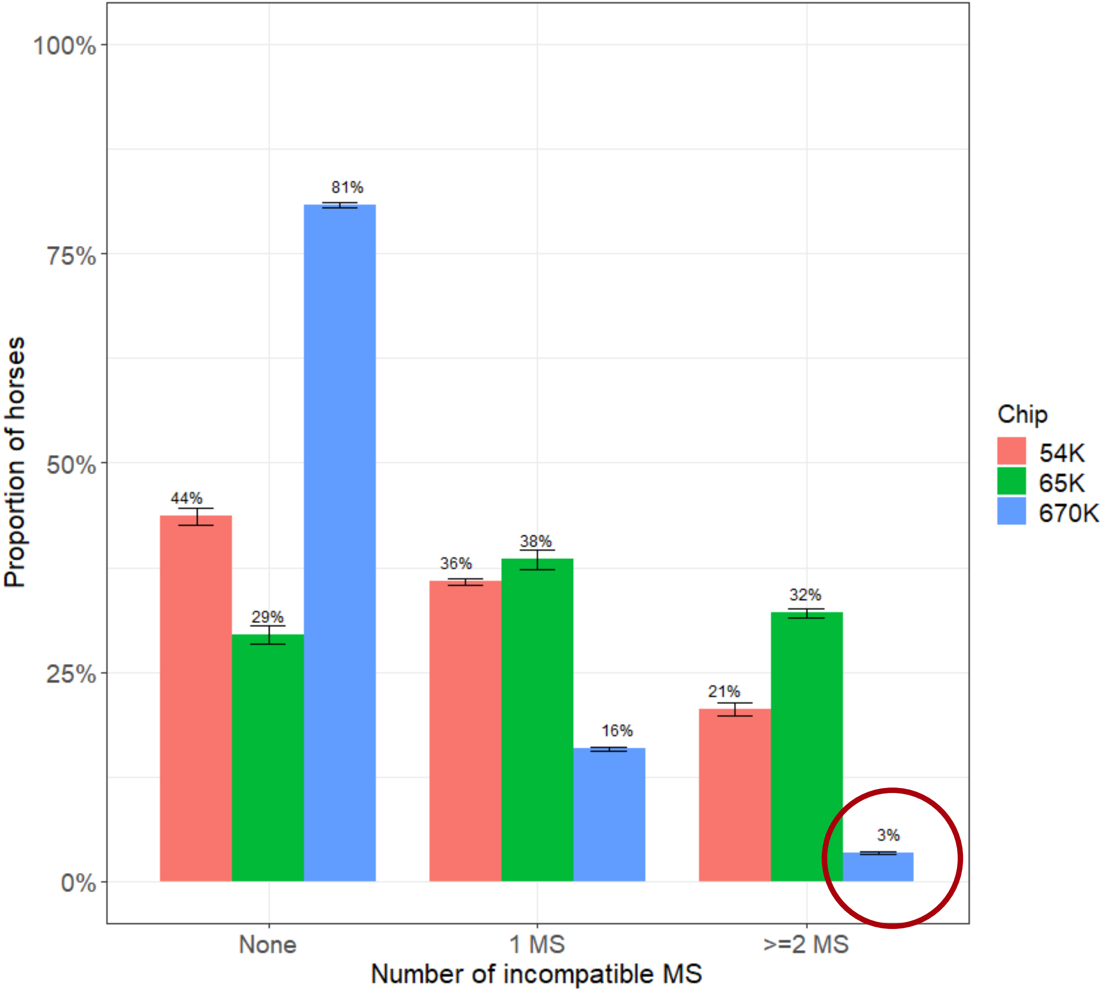
RESULTS

Effectiveness varied greatly with the density of the chip used

Horses genotyped with the 670K chip had only 3.4% of incompatibilities with more than one MS false, instead of 32.1% for the 65K

For every breed within the 670K chip we have observed good results for the incompatibilities: 4% for Arabs, 3.4% for Selle-Français and 3.2 for French Trotters

For further use: keep markers from the 670K chip they are certainly closer to the MS



Discussion

A correspondence table was built and there were only 3% of incompatibilities with more than 1 MS with SNPs from the 670K, closest to the MS → up to 97% re-genotyping costs saved

The limits of this study are :

- **This study was made only with 3 major breeds → more work required on other breeds (Thoroughbreds,…)**
- **Are we willing to take the 3% risk of incompatibility**
- **What about false positives ?**

The view of ISAG :

There are 2 risks using imputed microsatellites:

- **The imputed profile produces incorrect alleles and a pedigree is falsely qualified.**
- **The imputed profile produces incorrect alleles and the pedigree is falsely excluded, STR's need to be run to verify the imputed profile is correct/or not , so the exclusion can be reported or the offspring can be qualified.**

Acknowledgments and references

- **Thanks to:**

Fonds Éperon INRAE SF LeTROT plus qu'une allure, une culture Aca

REFERENCES:

- **McClure MC, Sonstegard TS, Wiggans GR, et al. Imputation of microsatellite alleles from dense SNP genotypes for parentage verification across multiple Bos taurus and Bos indicus breeds. *Front Genet.* 2013;4:176. Published 2013 Sep 18. doi:10.3389/fgene.2013.00176**
- **McClure M, Sonstegard T, Wiggans G, Van Tassell CP. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. *Front Genet.* 2012;3:140. Published 2012 Aug 14. doi:10.3389/fgene.2012.00140**
- **Sargolzaei M, Chesnais J.P, Schenkel F.S. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics.* 2014. doi: 10.1186/1471-2167-15-478**