

Comparison of genomic selection approaches for small breeds

C. Hozé^{1,2}, S. Fritz², D. Boichard¹, V. Ducrocq¹, P. Croiseau¹

1 INRA, UMR 1313, GABI, Jouy-en-Josas, France

2 UNCEIA, 149 rue de Bercy, Paris, France



Context

- **Genomic selection aims to predict breeding values (EBV) using DNA information**
- **Evaluations are accurate when reference population (RP) is large and are now implemented worldwide**
- **In France : implemented for the 3 main dairy breeds**
- **For small breeds : constitution of RP is challenging**
 - **Technically: low number of progeny tested bulls per year**
 - **Economically: proportionally higher genotyping cost**



Reference population size in France



Holstein : 24 772



Montbéliarde : 2616



Normande : 2344



Abondance : 209



Tarentaise : 185



Simmental : 125



Brown Swiss : 90



Vosgienne : 54



Pie Rouge : 51

In dairy cattle : From 50 to 25 000 bulls per breed



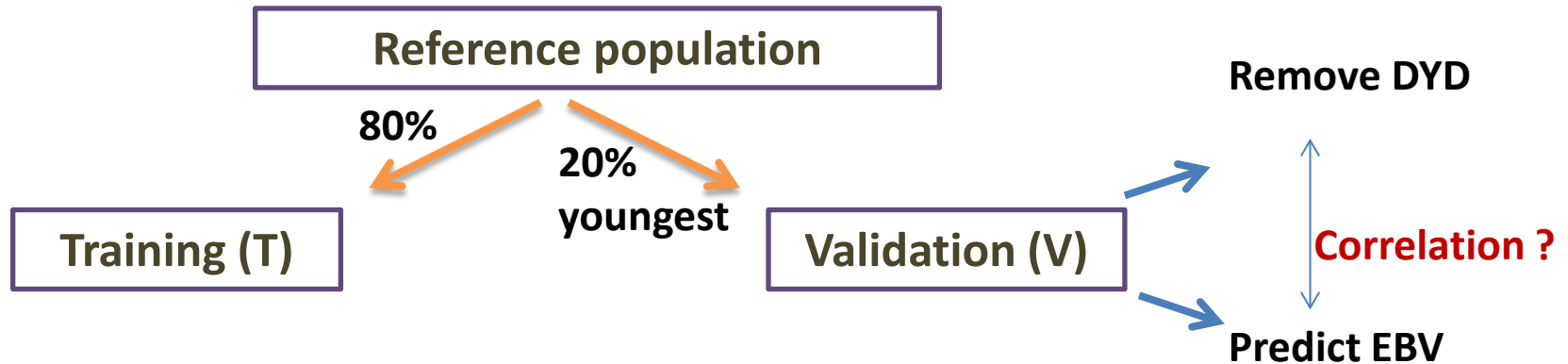
Context

- **How to implement GS with < 500 animals ?**
 - **Within breed**
 - **Across breeds**
 - **Pool RP from several breeds and develop multi-breed GS**
 - **Take advantage of short distance LD from HD-genotypes**
- **Three approaches tested here :**
 - 1) **Multi-breed GS for small breeds**
 - 2) **Within-breed GS for small breeds**
 - 3) **Comparison with GS results for large breeds**



Methods

- How to estimate accuracy of GS ?



- For small breeds : unstable results because of limited number of validation animals

- ➔ Use a large breed (Normande breed) to mimic small breeds
 - Reduce the training dataset but not the validation dataset



Method

- **Set up of a large HD reference population**

- 1) Genotype a population on the HD chip (« imputation pop »)
 - 2) Impute the reference population genotyped on the 50K chip
- Imputation accuracy > 99% (*Hoze et al., GSE 2013*)
 - Outcome : a large reference population with HD-genotypes

	Montbéliarde (MO)	Holstein (HO)	Normande (HO)
HD-genotyped	527	773	535
Bulls with daughters in France (EDC > 5 ; imputed + genotyped)	1788	4989	1991



Methods : three datasets studied

Validation set : 20% youngest bulls

Dataset **A** : All but validation set

Validation :

394 Normande
HD-genotyped
+ imputed

Training population A :

1597 Normande
HD-genotyped
+ imputed



Methods : three datasets studied

Validation set : 20% youngest bulls

Dataset B : Only HD-genotyped bulls of dataset A

Validation :

**394 Normande
HD-genotyped
+ imputed**

Training population B :

**404 Normande
HD-genotyped**



Methods : three datasets studied

Validation set : 20% youngest bulls

Dataset C : 50 % of dataset B (randomly chosen)

Validation :

394 Normande
HD-genotyped
+ imputed

Training population C :

194
Normande
50 % of HD-
genotyped



Methods : three datasets studied

For multi-breeds studies :

Set **A**, **B** or **C** pooled with bulls from Montbéliarde and Holstein

Validation :

394 Normande
HD-genotyped
+ imputed

Training population :

1597 Normande
HD-genotyped
+ imputed

194
Normande
50 % of HD-
genotyped

404
Normande
genotyped

1788 Montbéliarde
+ **4989 Holstein**



Methods : evaluations models

- **BLUP : pedigree based model**
- **BayesC (GS3 software *Legarra et al.*) with parameters :**
 - 1% of SNPs (≈ 7000) with a non-zero effect
 - 30% of genetic variance explained by residual polygenic
- **Bull's phenotypes :**
 - **DYD (Daughter yield deviation) standardized per breed :**
 - Derived from national genetic evaluation
 - Centered and divided by genetic standard deviation of the breed
 - **Trait studied**
 - Milk / Fat Yield (FY) / Protein Yield (PY)
 - Fat % / Protein % / Somatic Cell Score (SCS)



Results : Accuracy of prediction

Correlation and average regression slope between predicted EBV and DYD

Situation A : 194 T / 394 V

	Milk	FY	PY	Fat%	Pro%	SCS	Average corr	Average slope
BLUP	0.13	0.08	0.13	0.22	0.24	0.28	0.180	0.52
BayesCpi Within-breed GS	0.17	0.17	0.17	0.29	0.35	0.32	0.247	0.54
BayesCpi Multi-breed GS (+ 1788 MO +4989HO)	0.23	0.24	0.24	0.36	0.35	0.29	0.285	0.59

- + 6.7 % correlation with within-breed GS
- + 3.8 % correlation with multi-breed GS



Results : Accuracy of prediction

Correlation and average regression slope between predicted EBV and DYD

Situation B : 404 T / 394 V

	Milk	FY	PY	Fat%	Pro%	SCS	Average corr	Average slope
BLUP	0.23	0.30	0.24	0.31	0.38	0.40	0.310	0.66
BayesCpi Within-breed GS	0.31	0.39	0.33	0.40	0.47	0.42	0.387	0.73
BayesCpi Multi-breed GS (+ 1788 MO +4989HO)	0.35	0.40	0.35	0.48	0.49	0.43	0.416	0.71

- + 7.7 % correlation with within-breed GS
- + 2.9 % correlation with multi-breed GS



Results : Accuracy of prediction

Correlation and average regression slope between predicted EBV and DYD

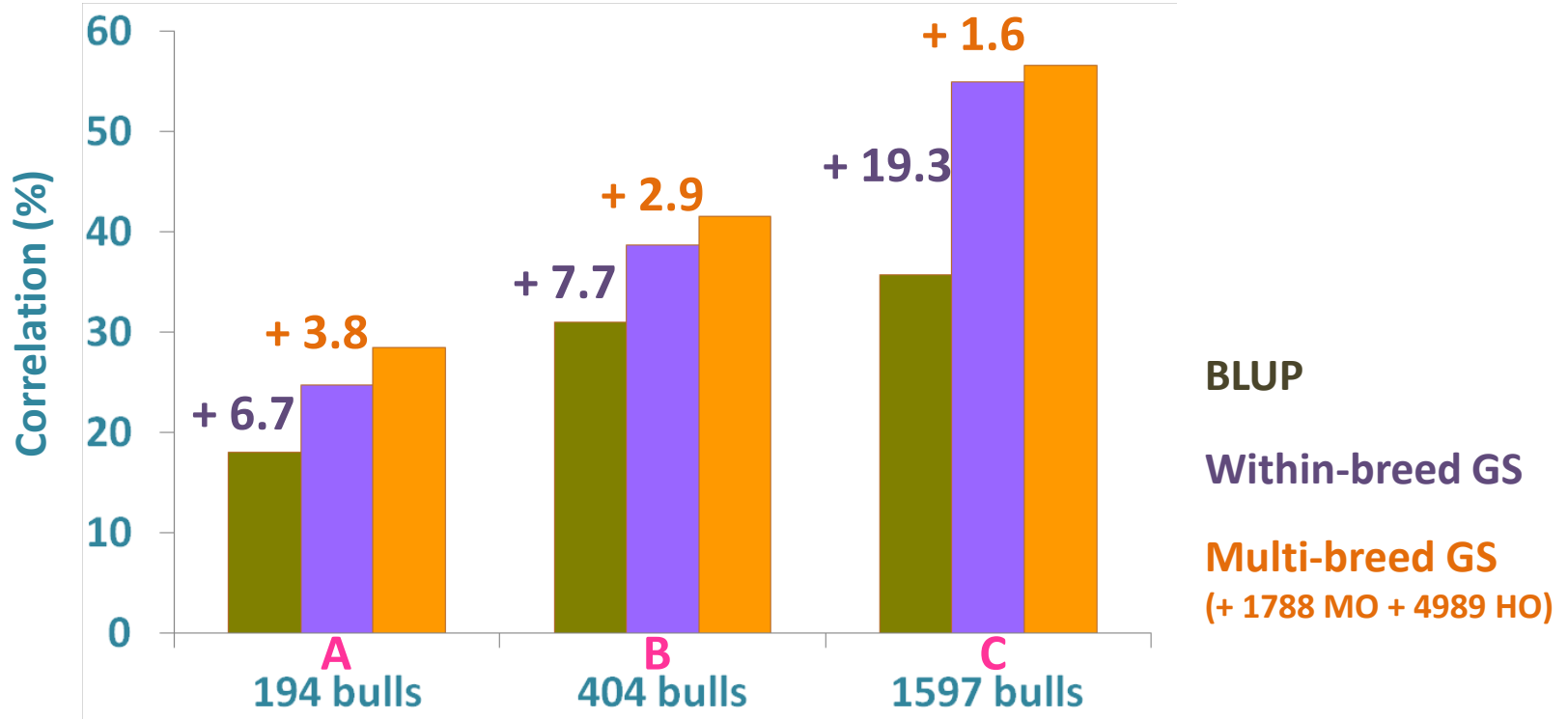
Situation C : 1597 T / 394 V

	Milk	FY	PY	Fat%	Pro%	SCS	Average corr	Average slope
BLUP	0.32	0.35	0.30	0.35	0.40	0.43	0.357	0.76
BayesCpi Within-breed GS	0.48	0.49	0.49	0.64	0.63	0.56	0.550	0.875
BayesCpi Multi-breed GS (+ 1788 MO + 4989 HO)	0.50	0.52	0.51	0.65	0.64	0.57	0.566	0.863

- + 19.3 % correlation with within-breed GS
- + 1.6 % correlation with multi-breed GS



Results : Impact of reference population size



- When increasing within breed reference population size :
 - Accuracy of GS prediction increases
 - Benefit of multi-breed GS declines



Conclusion: A first step for small breed GS

- **For all scenarios tested :**
 - Prediction of EBV is improved by GS
 - Accuracy increases with the number of animals in training population
- **Multi-breed GS provides an increased accuracy and is an appealing approach for small breeds**
- **Encouraging results but :**
 - Computational time is limiting (one week for 7000 animals)
 - Further developments are required



Acknowledgements



Thank you for your attention!

Any question ?



Chris HOZE

Le réseau de la génétique animale