



Successful genotype imputation from medium to high density in Belgian Warmblood horses



L. Chapard, R. Meyermans, W. Gorssen, B. Van Mol, F. Pille, N. Buys and S. Janssens

Center for Animal Breeding and Genetics, Department of Biosystems, KU Leuven, 3001 Leuven, Belgium

Department of Surgery and Anaesthesiology of Domestic Animals, Faculty of Veterinary Medicine, Ghent University, 9820 Merelbeke, Belgium

Session 90: "Genetics in horses"



KU LEUVEN



SNPs and advantages of genomics

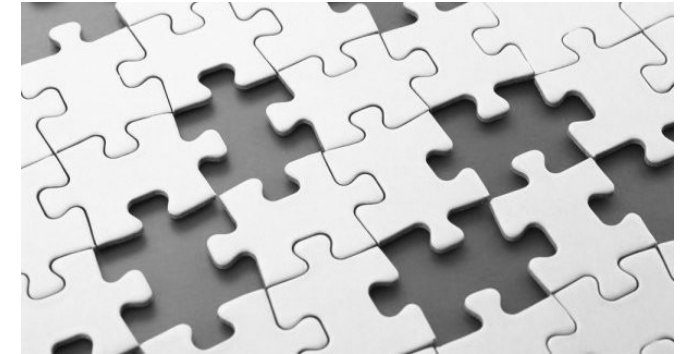
- **SNP** (Single nucleotide polymorphism)
- Medium density (70K) and high density (670K) **SNP arrays are commercially available** in horse breeding
- **Genotype data** permits to do **genomic selection, GWAS, parentage analysis, genetic diversity studies, etc...**





Genotype imputation

- **Genotype imputation:**
 - Increases the **density of available genotypes within a population**
 - Showed **great results in livestock species** from medium (MD) to high (HD) density genotypes
 - Can be performed on closely related breeds

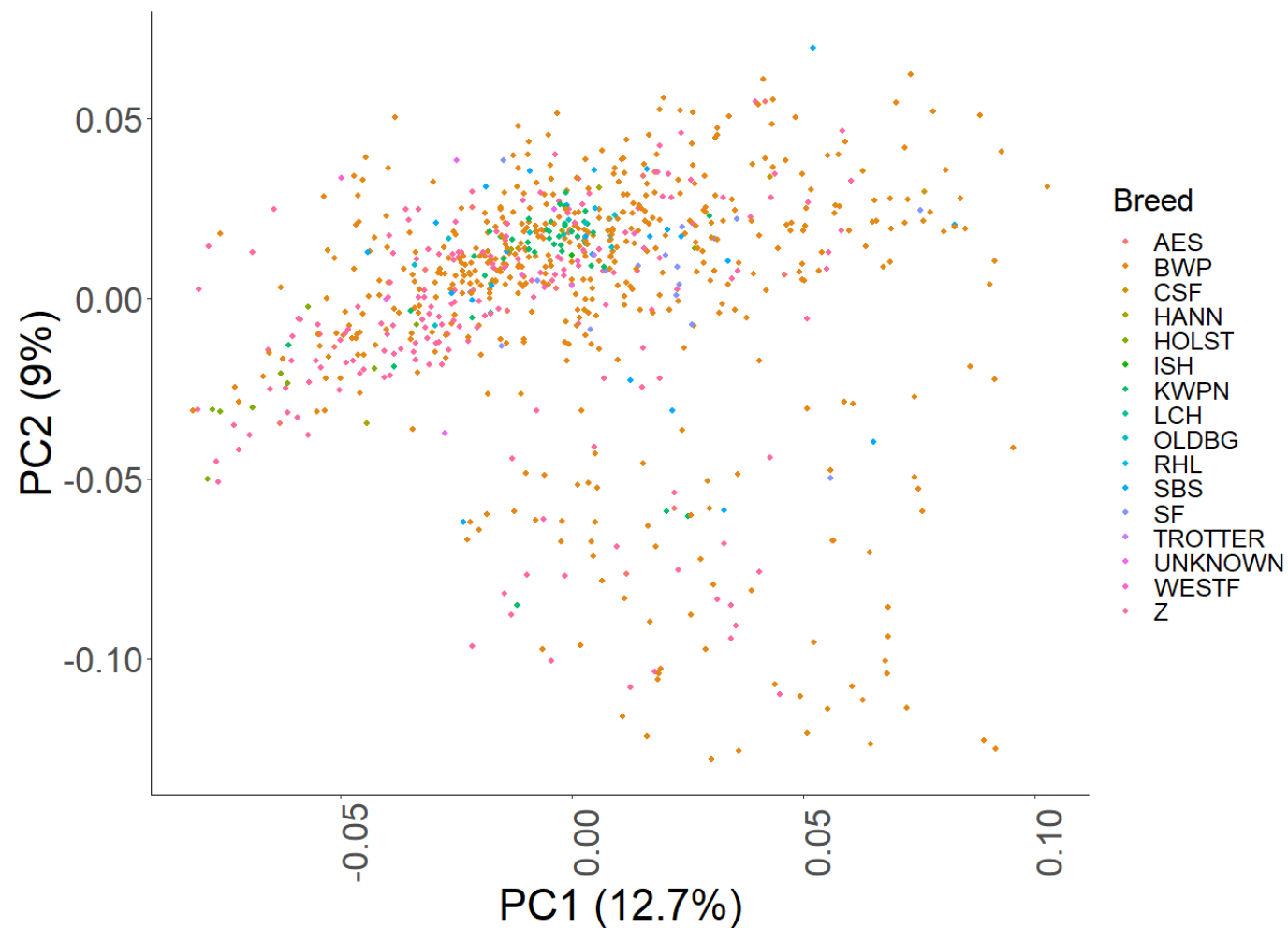


Process of inferring unobserved genotypes in a sample of individuals using a reference panel



Genotype data

- 823 HD genotypes of sport horses:



Breed	n
BWP	490
Z	189
KWPN	41
SBS	24
SF	22
AES	14
HOLST	11
OLDBG	11
WESTF	6
HANN	5
ISH	1
LCH	1
RHL	1
TROTTER	1
UNKNOWN	6



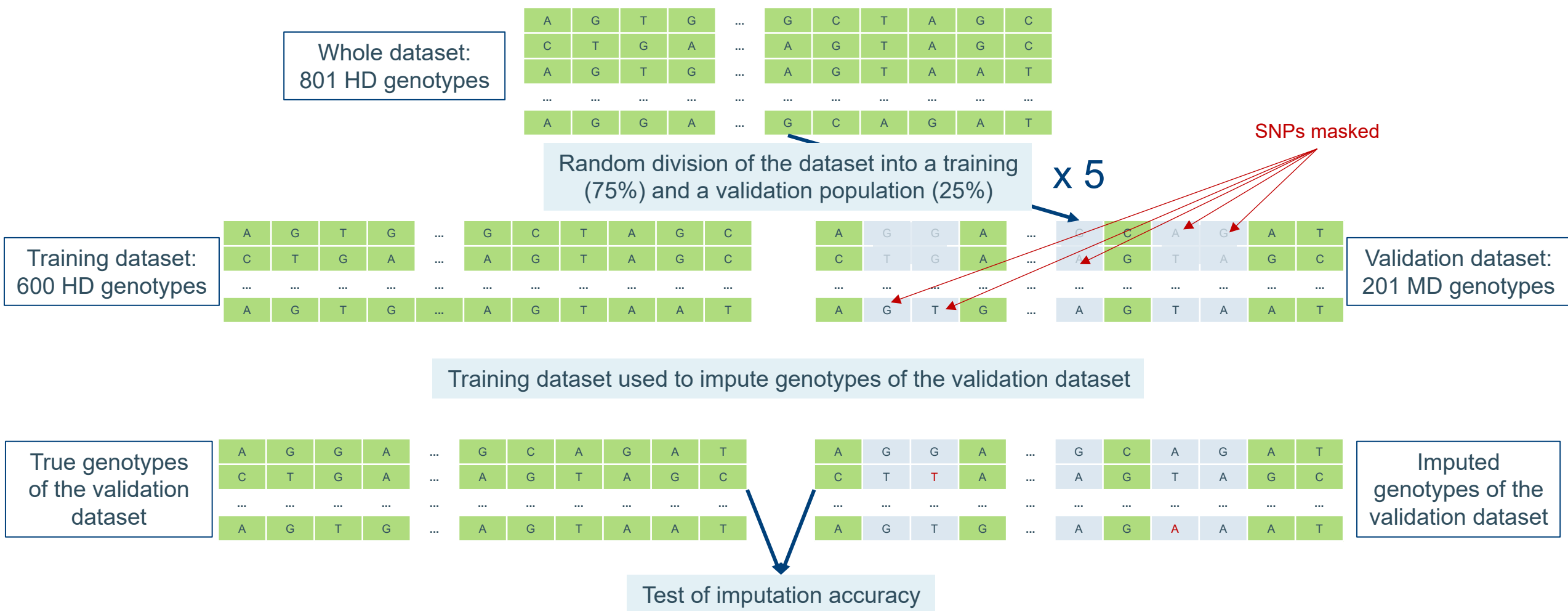
Methods

- **Quality control steps** (using Plink):
 - SNP positions updated from EquCab 2.0 to EquCab 3.0
 - Merge 4 datasets (248 horses genotyped in 2017, 192 in 2019, 192 in 2022 and 191 in 2023)
 - Keep only autosomes
 - Remove individuals with call-rate ≤ 0.95 , outlying heterozygosity $\geq$ mean heterozygosity ± 3 sd and duplicated individuals
 - Check population substructure via PCA
 - Remove SNPs with call-rate ≤ 0.95
- **55 509 SNPs in common between Affymetrix Axiom Equine 670K and GGP Equine 70K arrays**
- **Cross-validation scheme:** HD dataset divided into a reference panel (75% of the dataset) and a validation set (25% of the dataset) (**X5**)
- **Haplotype phasing** and **genotype imputation** using Beagle
- **Imputation accuracy** checked for each validating set using R

	n before QC	n after QC
Horses	823	801
SNPs	670 806	513 223



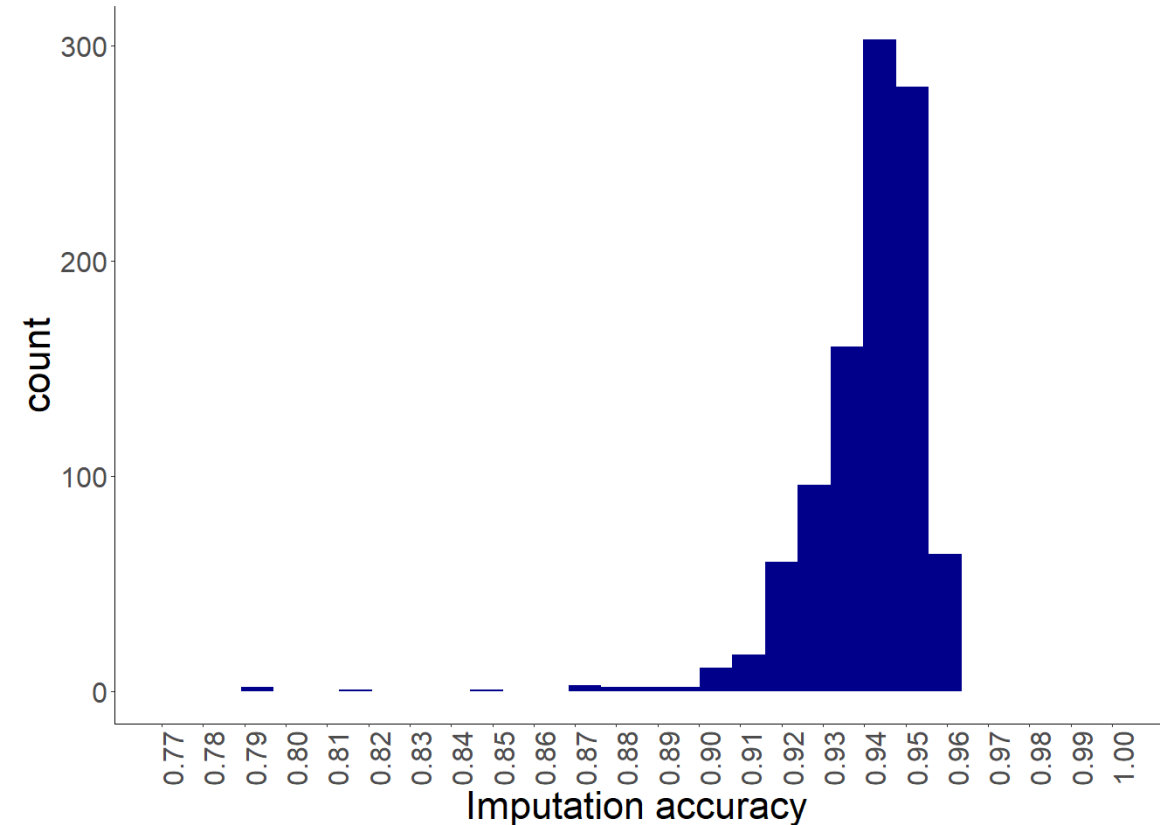
Cross-validation scheme





Genotype imputation accuracy per individual

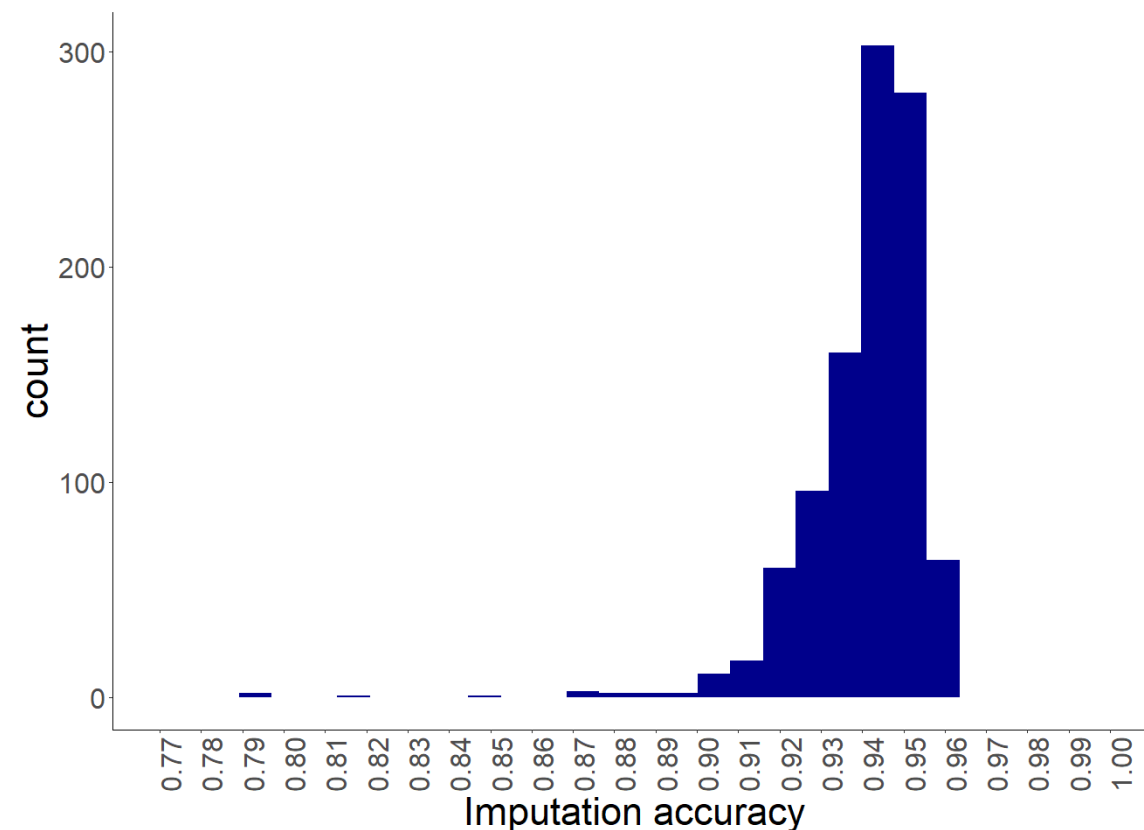
	min	mean	max	SD
Validating set 1	0.79	0.94	0.96	0.02
Validating set 2	0.87	0.94	0.96	0.01
Validating set 3	0.85	0.94	0.96	0.02
Validating set 4	0.89	0.94	0.96	0.01
Validating set 5	0.79	0.94	0.96	0.02





Genotype imputation accuracy per individual

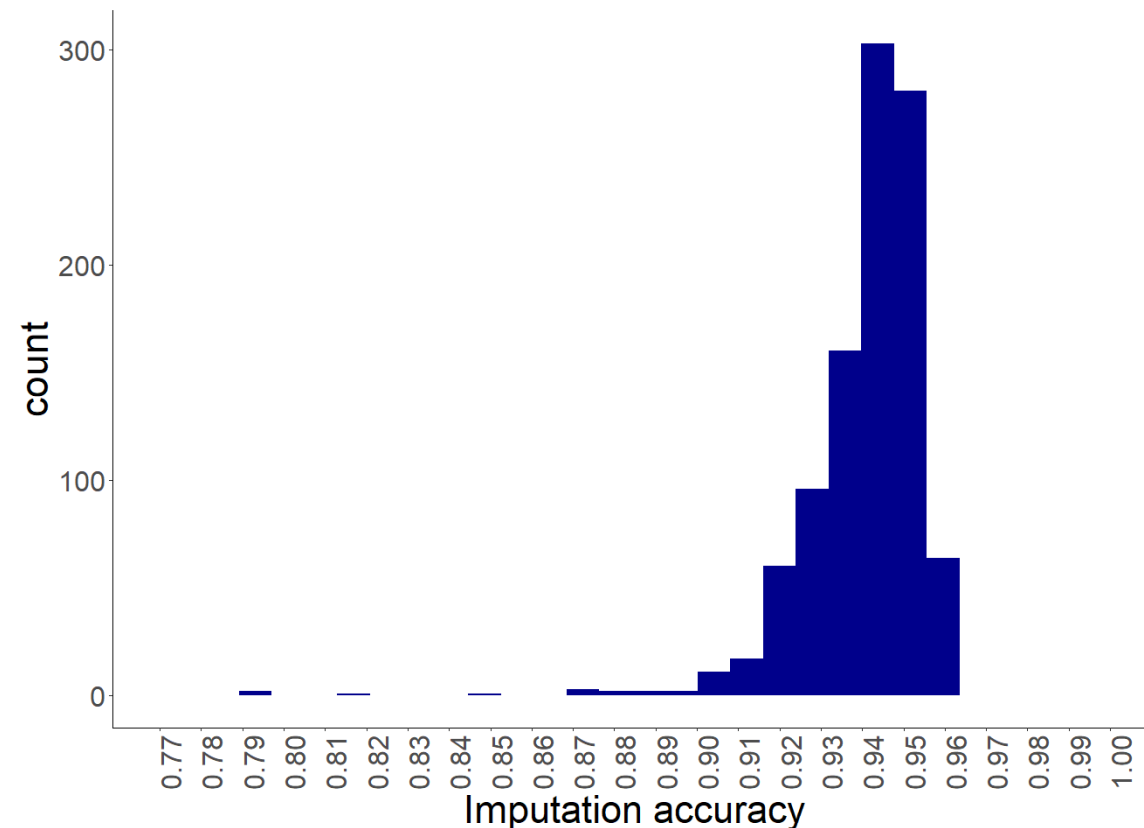
	min	mean	max	SD
Validating set 1	0.79	0.94	0.96	0.02
Validating set 2	0.87	0.94	0.96	0.01
Validating set 3	0.85	0.94	0.96	0.02
Validating set 4	0.89	0.94	0.96	0.01
Validating set 5	0.79	0.94	0.96	0.02





Genotype imputation accuracy per individual

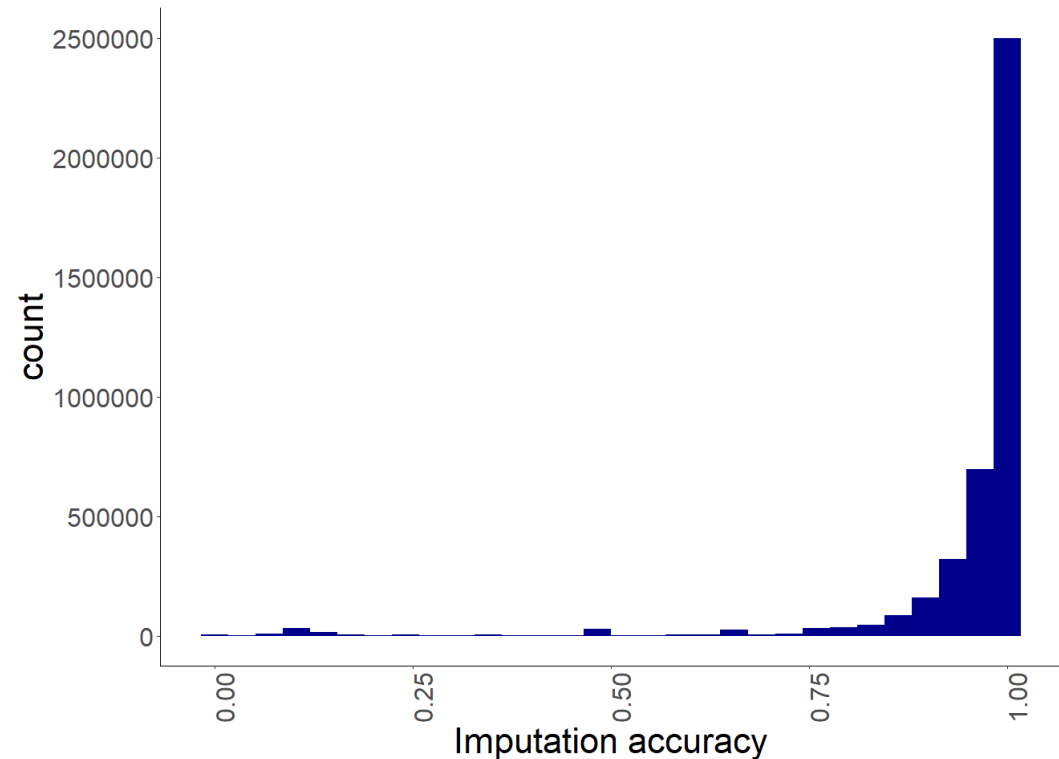
	min	mean	max	SD
Validating set 1	0.79	0.94	0.96	0.02
Validating set 2	0.87	0.94	0.96	0.01
Validating set 3	0.85	0.94	0.96	0.02
Validating set 4	0.89	0.94	0.96	0.01
Validating set 5	0.79	0.94	0.96	0.02





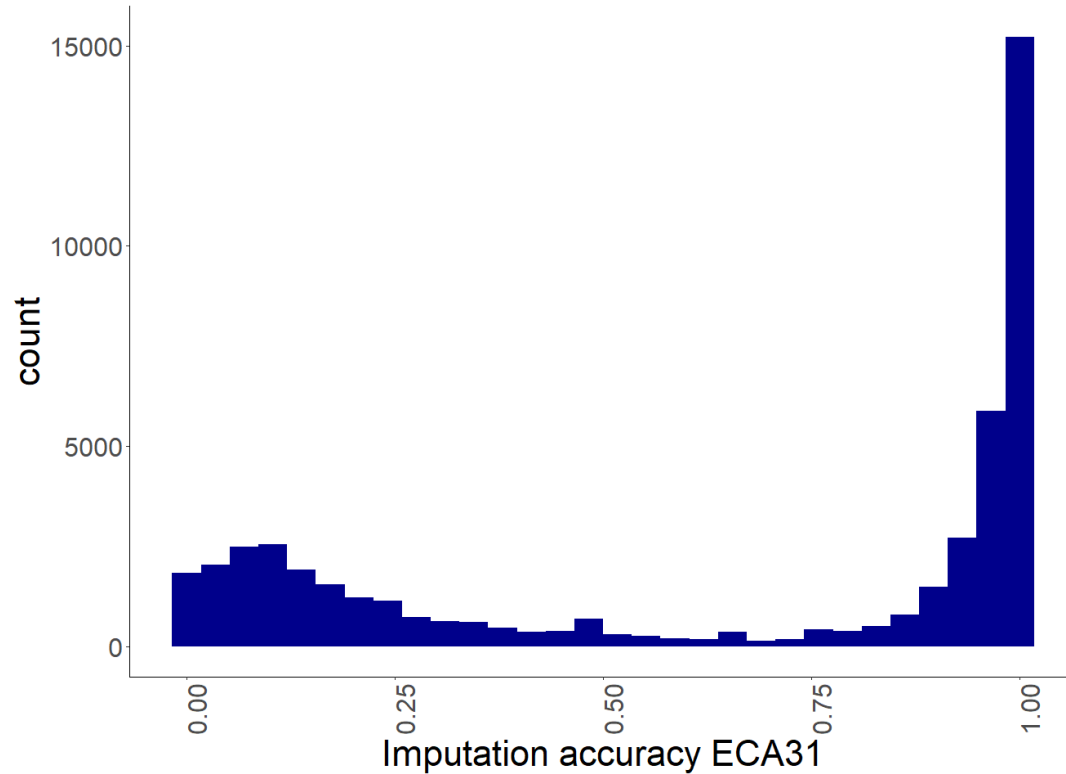
Genotype imputation accuracy per SNP

	min	mean	max	SD
Validating set 1	0	0.95	1	0.15
Validating set 2	0	0.94	1	0.15
Validating set 3	0	0.95	1	0.15
Validating set 4	0	0.95	1	0.15
Validating set 5	0	0.95	1	0.15





Genotype imputation accuracy per SNP

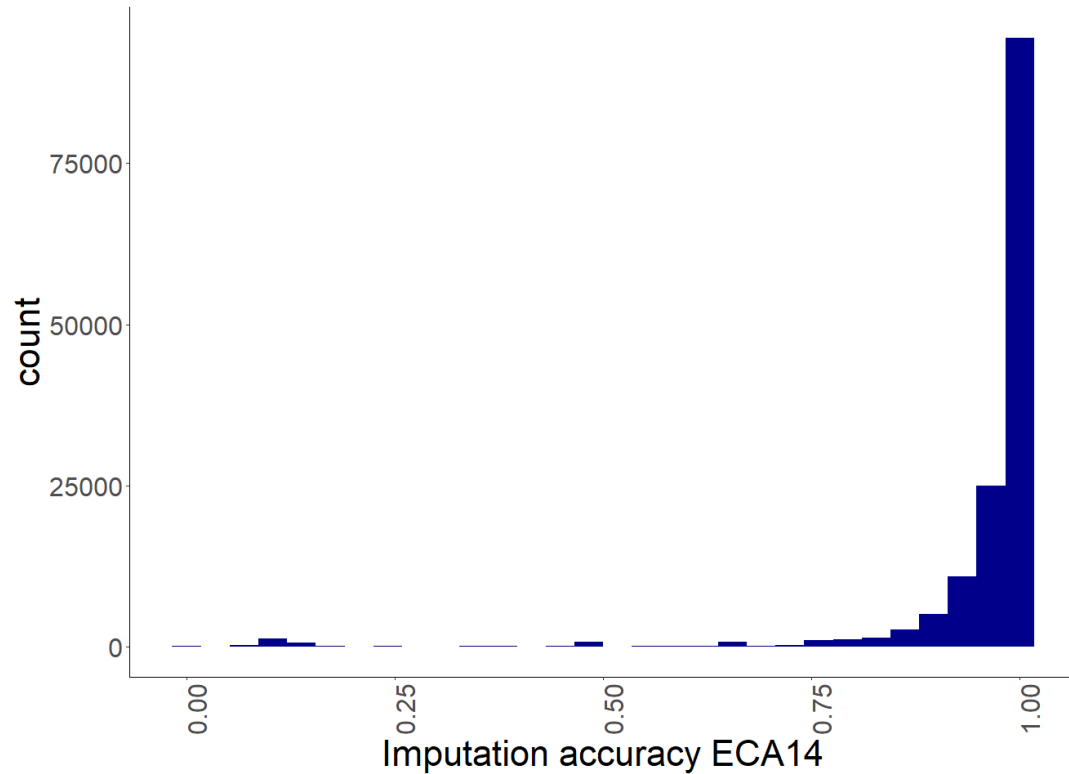


chr	mean r ²
ECA1	0.95
ECA2	0.95
ECA3	0.95
ECA4	0.95
ECA5	0.96
ECA6	0.95
ECA7	0.95
ECA8	0.95
ECA9	0.95
ECA10	0.95
ECA11	0.95
ECA12	0.92
ECA13	0.84
ECA14	0.96
ECA15	0.96

chr	mean r ²
ECA16	0.95
ECA17	0.95
ECA18	0.95
ECA19	0.94
ECA20	0.95
ECA21	0.95
ECA22	0.95
ECA23	0.95
ECA24	0.95
ECA25	0.95
ECA26	0.95
ECA27	0.95
ECA28	0.95
ECA29	0.95
ECA30	0.95
ECA31	0.64



Genotype imputation accuracy per SNP



chr	mean r ²
ECA1	0.95
ECA2	0.95
ECA3	0.95
ECA4	0.95
ECA5	0.96
ECA6	0.95
ECA7	0.95
ECA8	0.95
ECA9	0.95
ECA10	0.95
ECA11	0.95
ECA12	0.92
ECA13	0.84
ECA14	0.96
ECA15	0.96

chr	mean r ²
ECA16	0.95
ECA17	0.95
ECA18	0.95
ECA19	0.94
ECA20	0.95
ECA21	0.95
ECA22	0.95
ECA23	0.95
ECA24	0.95
ECA25	0.95
ECA26	0.95
ECA27	0.95
ECA28	0.95
ECA29	0.95
ECA30	0.95
ECA31	0.64



Conclusion

- **Genotype imputation** from MD to HD genotypes showed **high accuracies**:
 - Per individual: **0.94** on average
 - Per SNP: **0.95** on average
- **Lowest accuracy for ECA31, highest for ECA5, ECA14, ECA15**

⇒ This panel can be used as a **reference panel** for **genotype imputation from MD to HD**



Next steps:

- **Use routinely** this **genotype imputation** from MD to HD in **Belgian Warmblood horses**
- **Develop** genotype imputation for **other Belgian horse populations**



Acknowledgments

- Data providers



- Funding



DEPARTEMENT
LANDBOUW
& VISSERIJ



Thank you for your attention



lea.chapard@kuleuven.be