From the 50K chip to the HD: Imputation efficiency in 9 dairy cattle breeds

Chris HOZÉ

INRA, GABI

UNCEIA

le réseau de la génétique animale

Genomic evaluations

- Require large reference population
- Implemented in the three French main dairy breeds
  - Holstein, Normande, Montbéliarde
  - Since 2008
  - Based on a 50K chip
- For other breeds ?
  - Problem : constitution of reference population
  - Solution : sharing reference population
    - Across countries? e.g. : Intergenomics for Brown Swiss
    - Across breeds ? e.g. : Gembal project

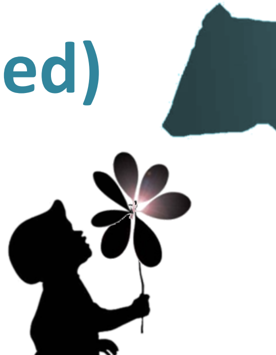
The GEMBAL project
Multibreed Genomics of Beef and Dairy Cattle

- **Develop multi-breed genomic selection**
  - **Share reference population across breeds**
  - **Develop genomic selection in all French cattle breeds**

- **Detect conserved chromosome segments accross breeds**
  - **Distance between markers < 10 Kb to maintain LD**
    - **50K chip : 54.609 SNP -> 1 SNP every 70 Kb** (BovineSNP50 BeadChip®)
    - **HD chip : 777.962 SNP -> 1 SNP every 4 Kb** (BovineHD BeadChip®)
  - **Possible on the high density chip ?**

The GEMBAL project
Multibreed Genomics of Beef and Dairy Cattle

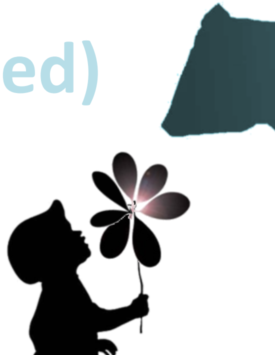
- **Genotype animals on the HD chip**
- **Study imputation efficiency from the 50K chip to the HD chip**
- **Impute 50K genotypes from national databases**
- **Study multi-breed GS using a large reference population with HD-genotypes (true and imputed)**

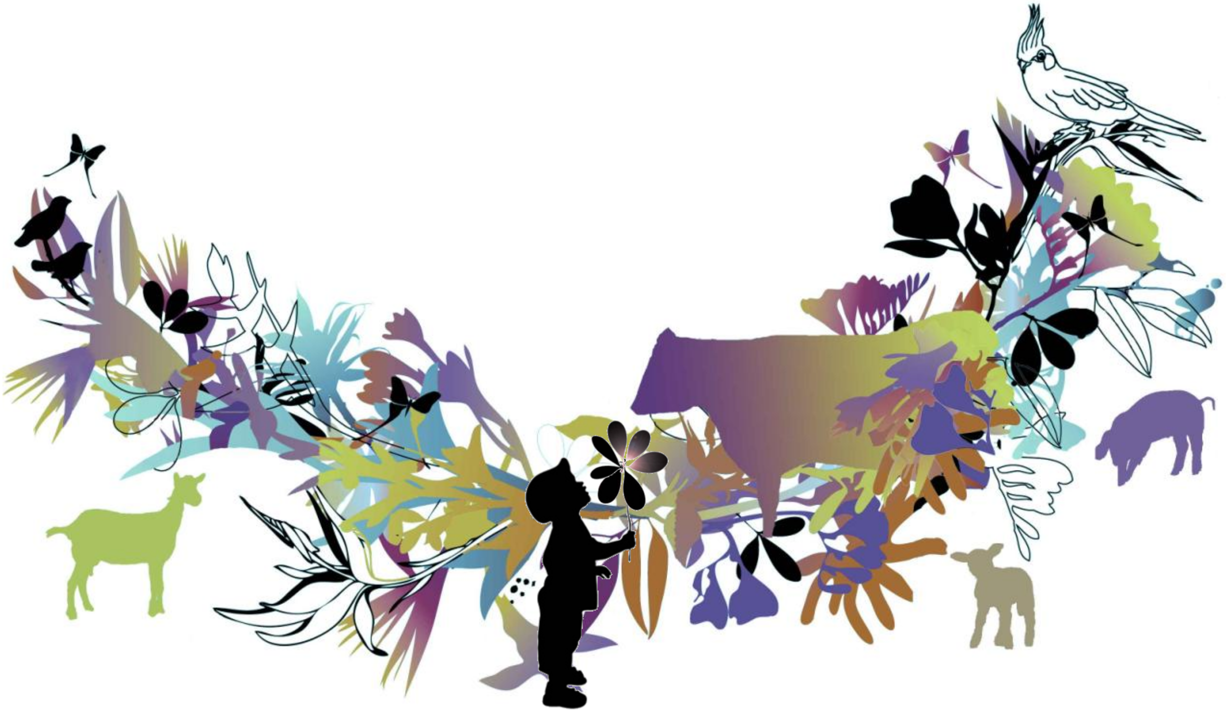


The GEMBAL project
Multibreed Genomics of Beef and Dairy Cattle

- Genotype animals on the HD chip
- **Study imputation efficiency from the 50K chip to the HD chip**
- Impute 50K genotypes from national databases
- Study multi-breed GS using a large reference population with HD-genotypes (true and imputed)

INSTITUT DE L'ELEVAGE





MATERIALS AND METHODS

Number of animals genotyped on the HD chip

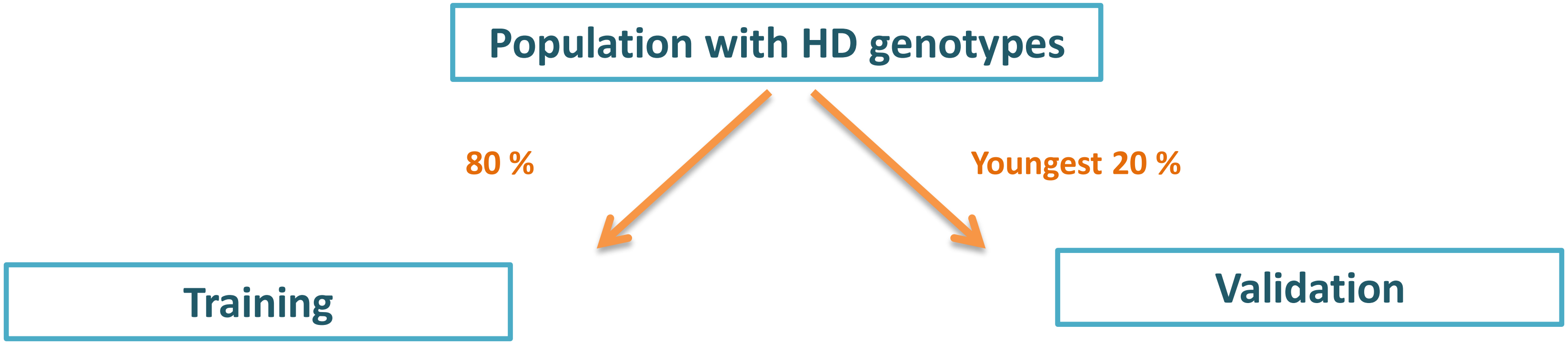
Breed Project / Consortium	Abondance	Brown Swiss	Tarentaise	Simmental	Montbéliarde	Normande	Vosgienne	Holstein	Pie Rouge (Red Pied cattle)
GEMBAL	209	99	184	124	522	546	52	236	41
EUROGENOMICS								540	9
Total	**209**	**99**	**184**	**124**	**522**	**546**	**52**	**776**	**50**

- **2,562 animals from 9 breeds**
- **From 50 to 776 animals per breed**

Quality control

- **Animals :**
  - Call rate above 0.95
  - Consistency between pedigree and SNP-based relationships
- **SNP :**
  - Criteria
    - Located on the 29 autosomes with UMD3 position
    - Call rate > 90%
    - p Hardy Weinberg > 10^{-4}
    - Consistent between the two chips
  - HD chip 777.962 SNP -> 664.808 SNP
  - 50K chip 54.609 -> 43.801 SNP

Computation of imputation efficiency



Population with HD genotypes

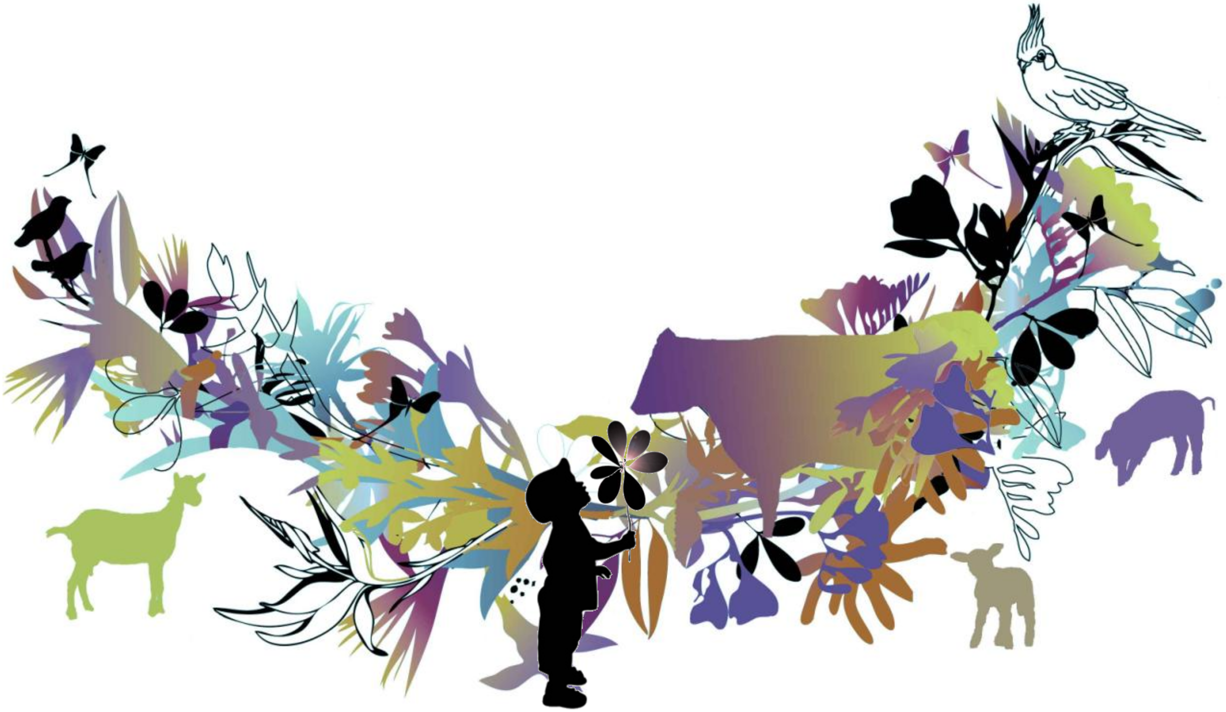
80 %

Youngest 20 %

Training

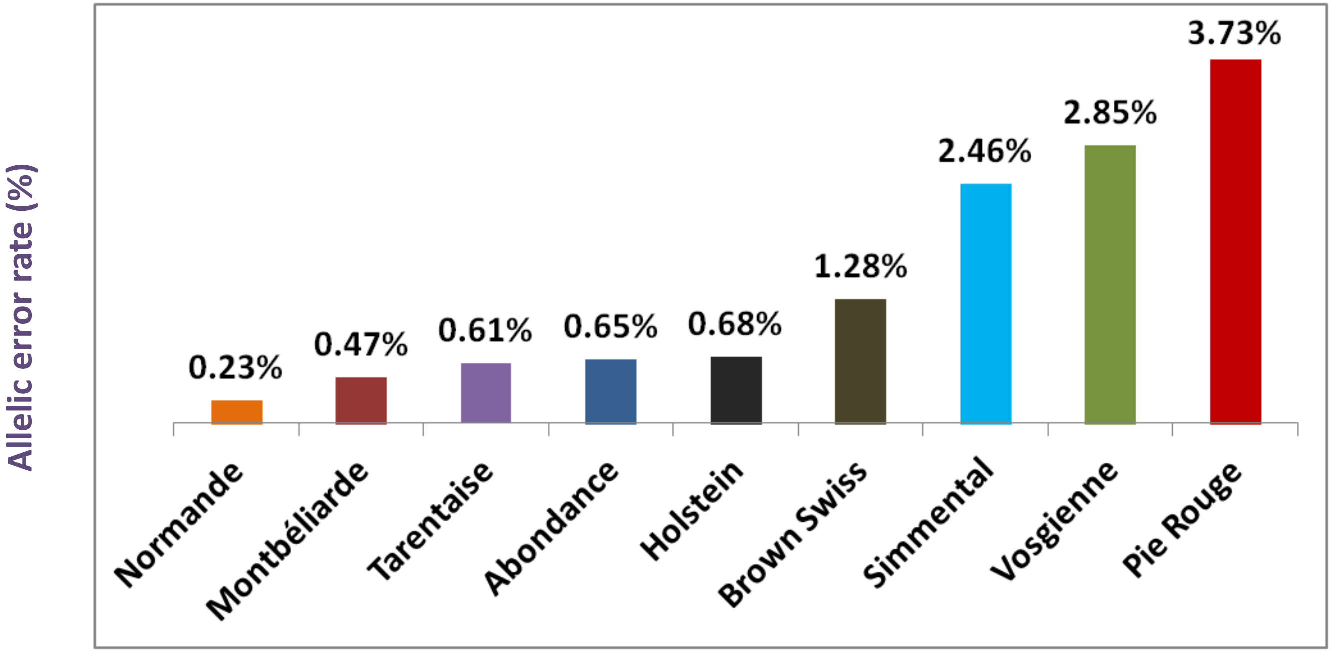
Validation

- For validation population
  - Erase markers specific to the HD chip
  - Impute them using Beagle 3.3.0 software
  - Compare true and imputed genotypes
  - Compute allelic imputation error rate



RESULTS

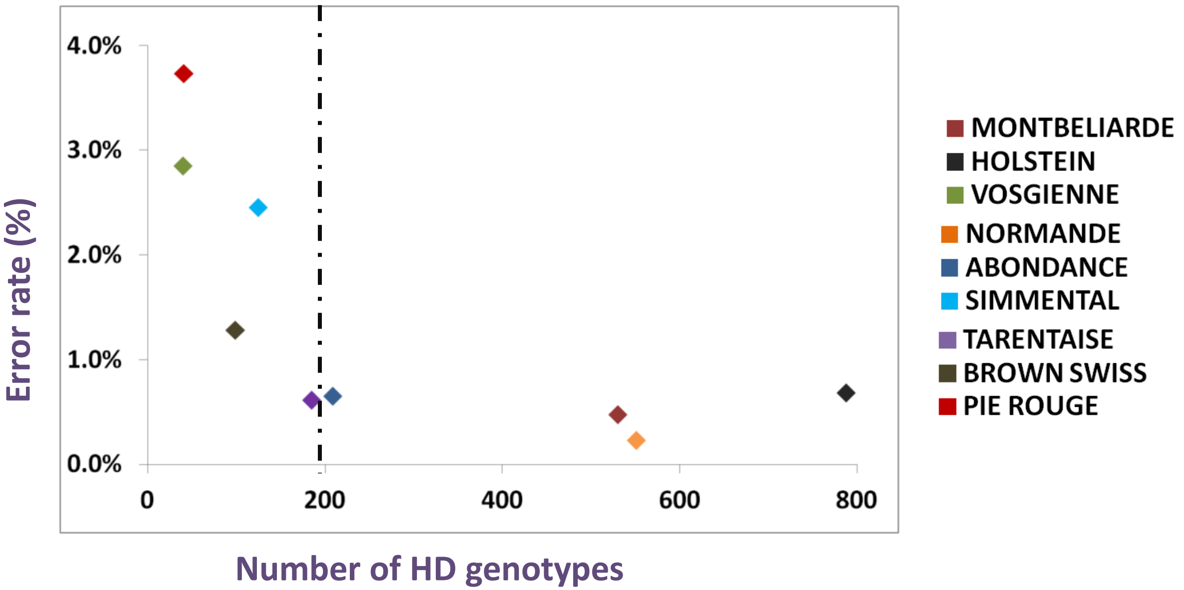
Imputation error rate per breed



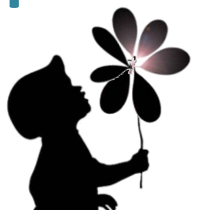
- **Efficient in large and medium size dairy breeds**
- **Higher error rates in foreign breeds and in small French cattle breeds**

Factors affecting imputation error rate

REFERENCE POPULATION SIZE

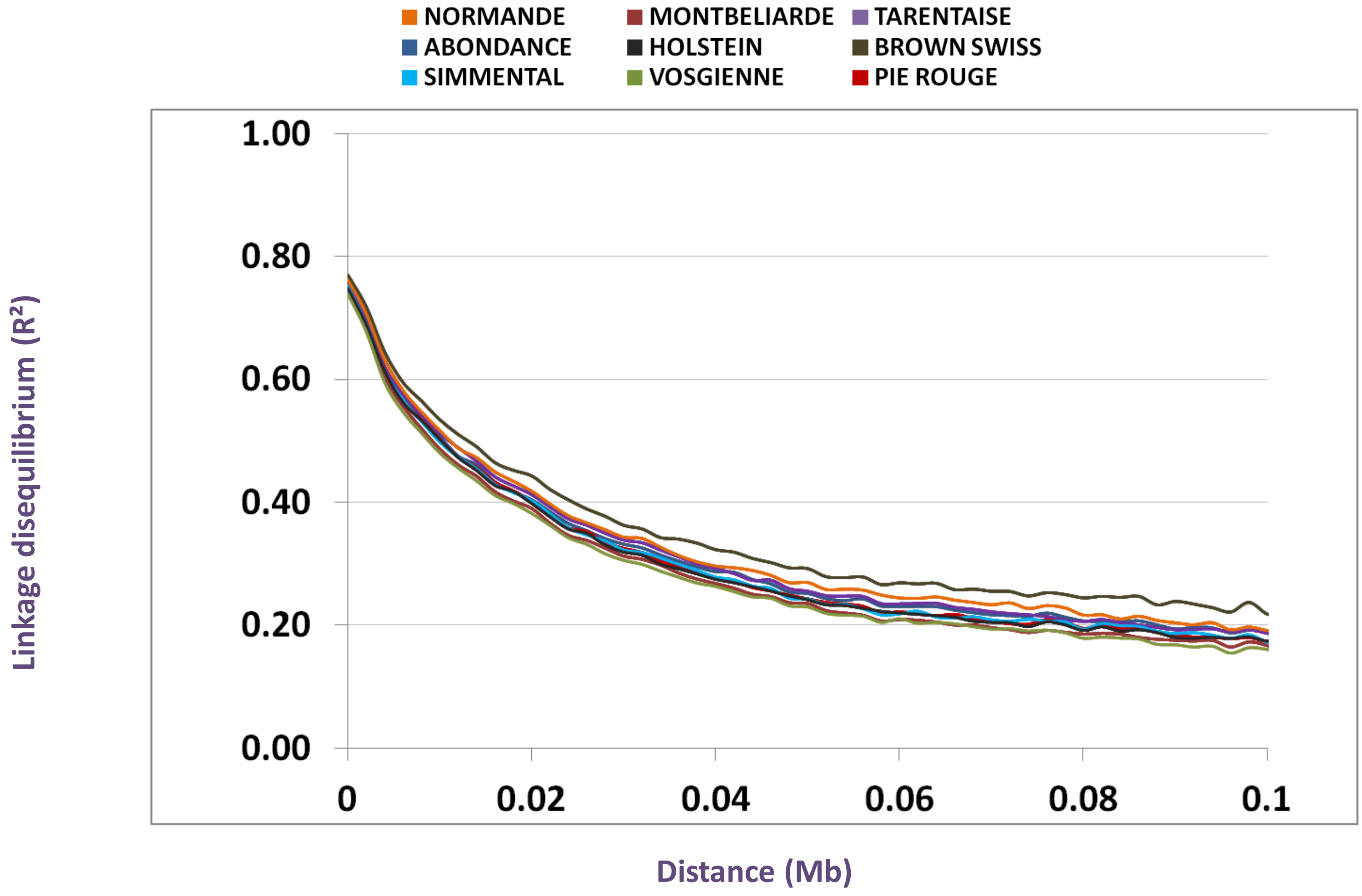


- **Strongly related to reference population size**
- **Efficient when more than 200 animals are genotyped**



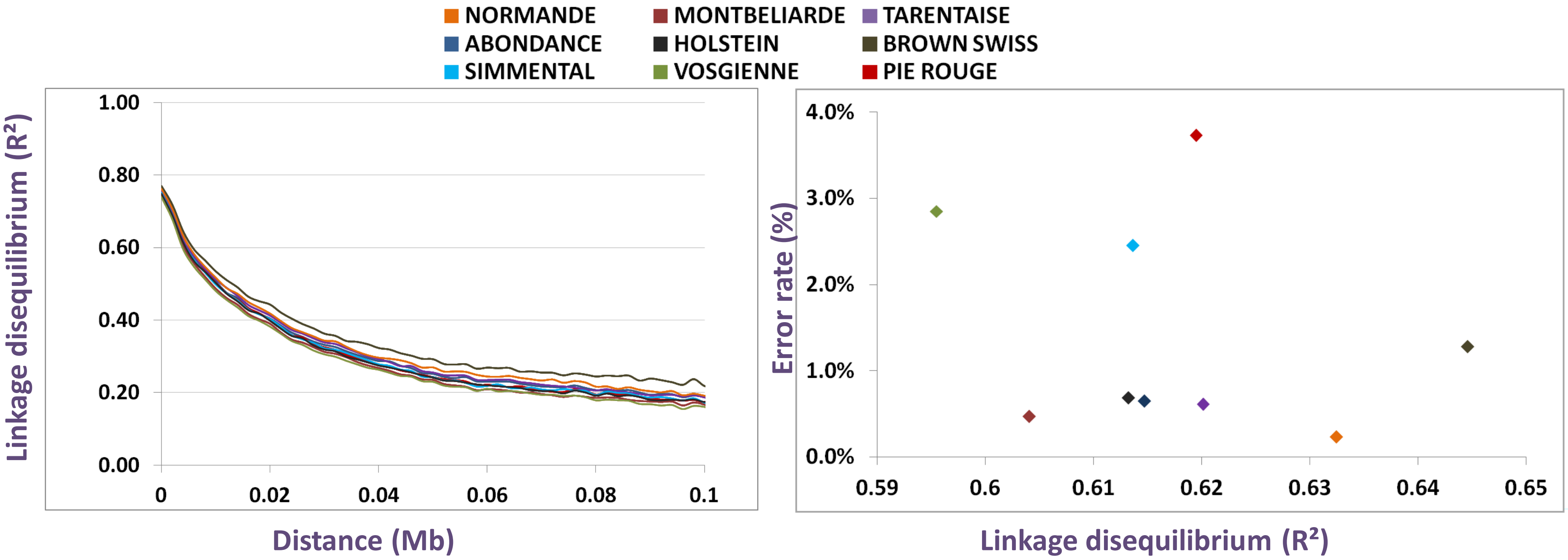
Factors affecting imputation error rate

LINKAGE DISEQUILIBRIUM DECAY



Factors affecting imputation error rate

LINKAGE DISEQUILIBRIUM DECAY **ERROR RATE AND LD LEVEL AT 4Kb**



- **No clear impact of LD level on imputation error rate**

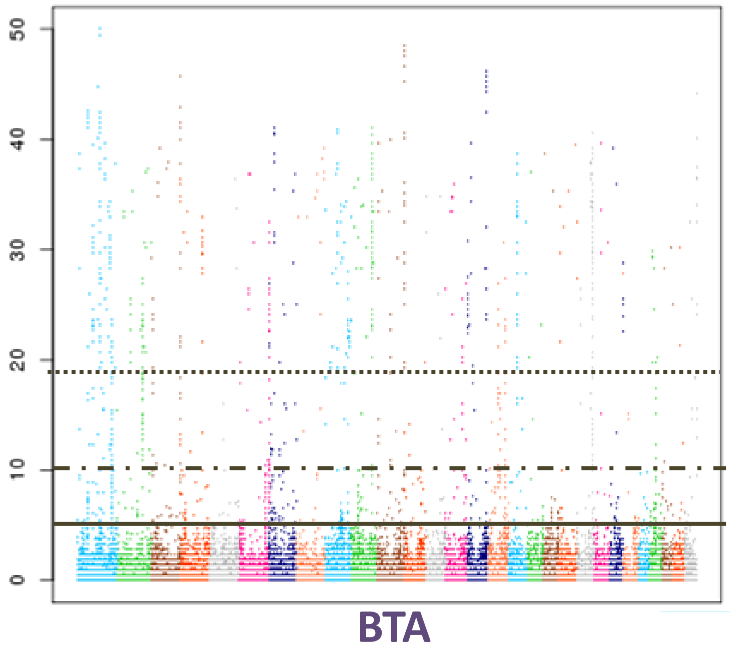
Imputation error rate in Montbéliarde

- **522 animals (416 T + 106 V)**
- **Mean error rate = 0.47%**

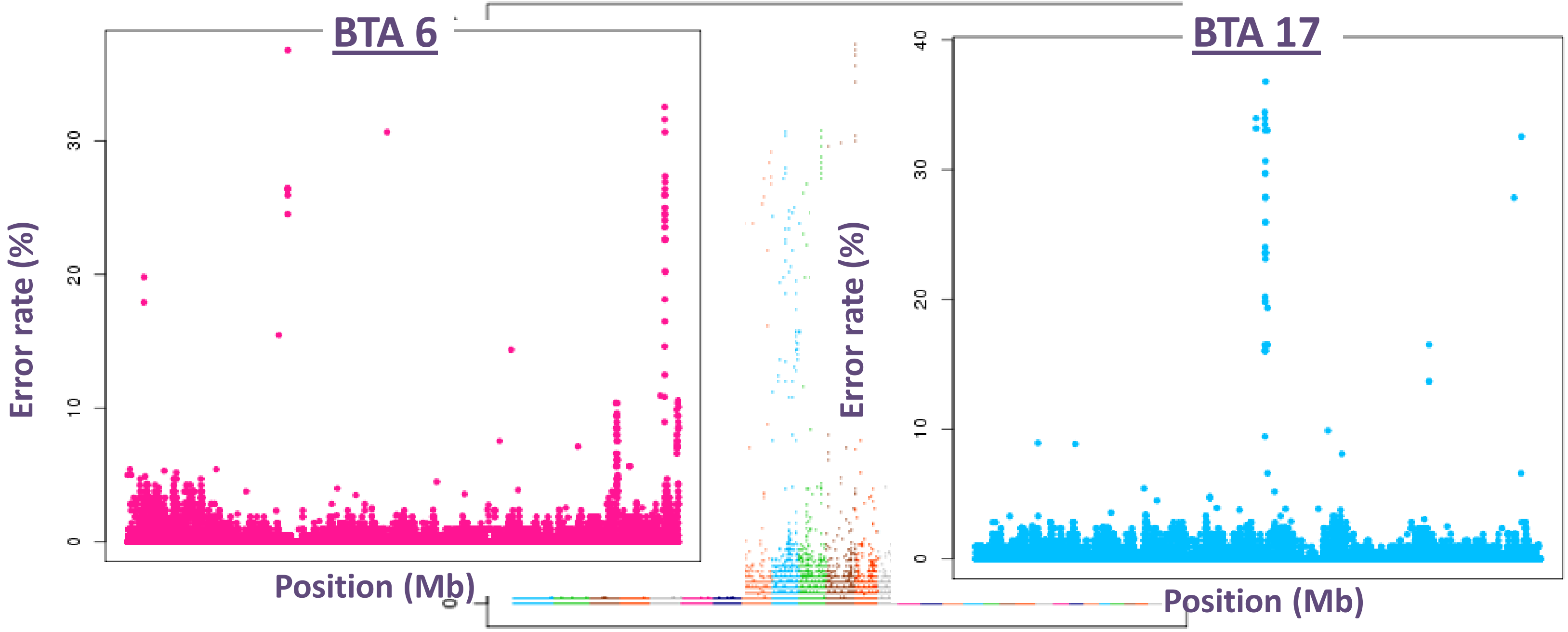
Imputation error rate in Montbéliarde

- **522 animals (416 T + 106 V)**
- **Mean error rate = 0.47%**
  - **Strong differences between SNP**
    - **13,104 SNPs with error rate > 5 %**
    - **8,214 SNPs with error rate > 10 %**
    - **6,030 SNPs with error rate > 20 %**

IMPUTATION ERROR RATE ALONG GENOME



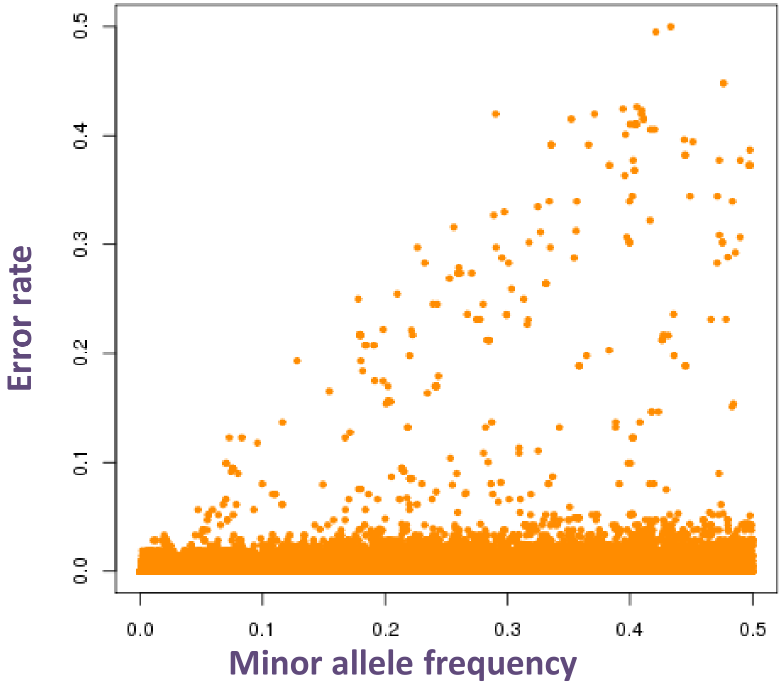
Imputation error rate in Montbéliarde



- **Higher error rate in chromosome extremities**
- **Some regions are likely to be mismapped**

Imputation error rate in Montbéliarde

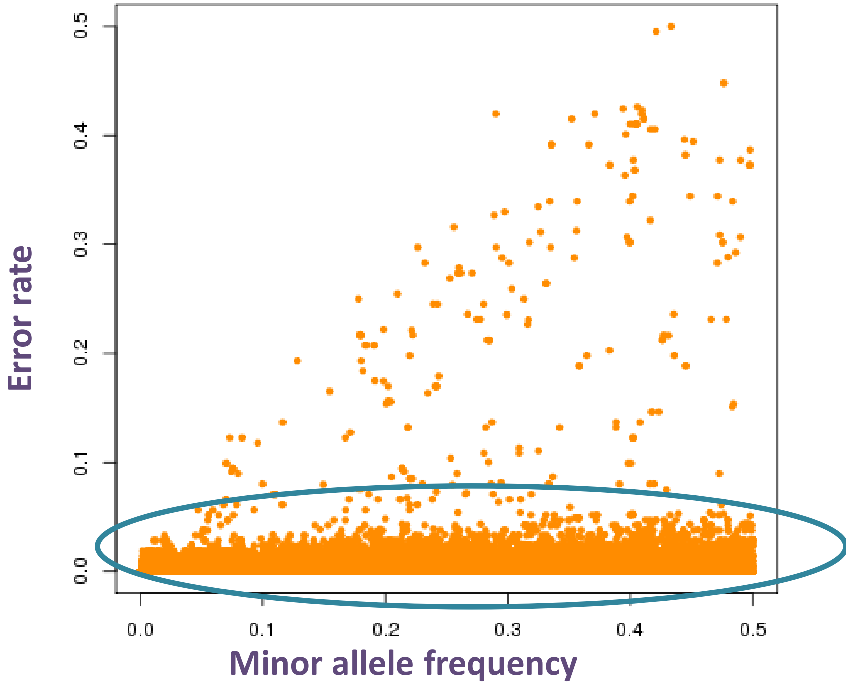
IMPUTATION ERROR RATE AND MAF



Imputation error rate in Montbéliarde

- **Almost no link for low error rate markers**

IMPUTATION ERROR RATE AND MAF

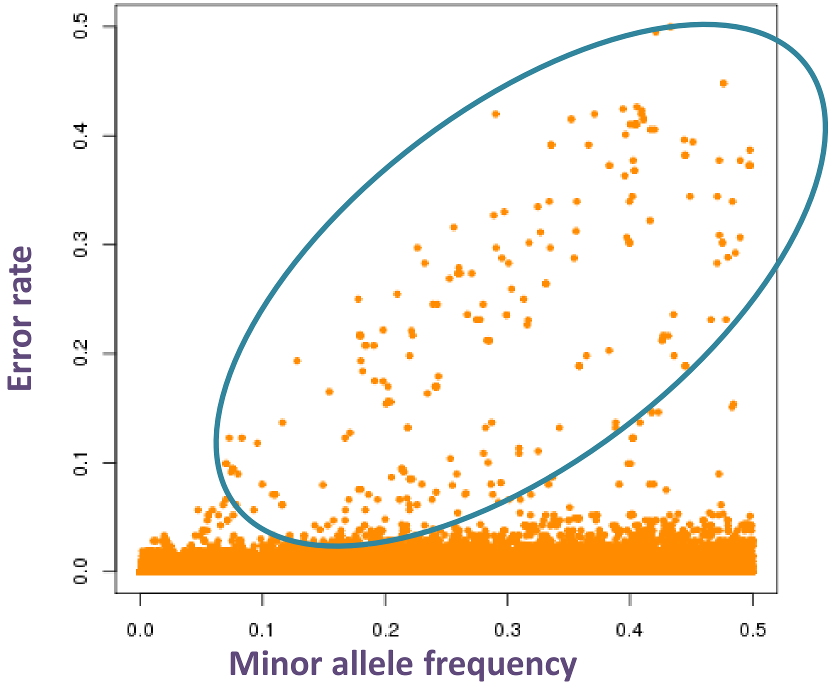


Imputation error rate in Montbéliarde

- **Almost no link for low error rate markers**

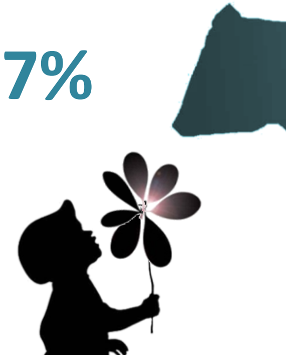
- **Correlation for high error rates markers**
  - **Due to SNP localization?**
  - **No LD between markers**
  - **Imputation = Random with respect to allele frequencies**
  - **Error rate ≈ MAF**

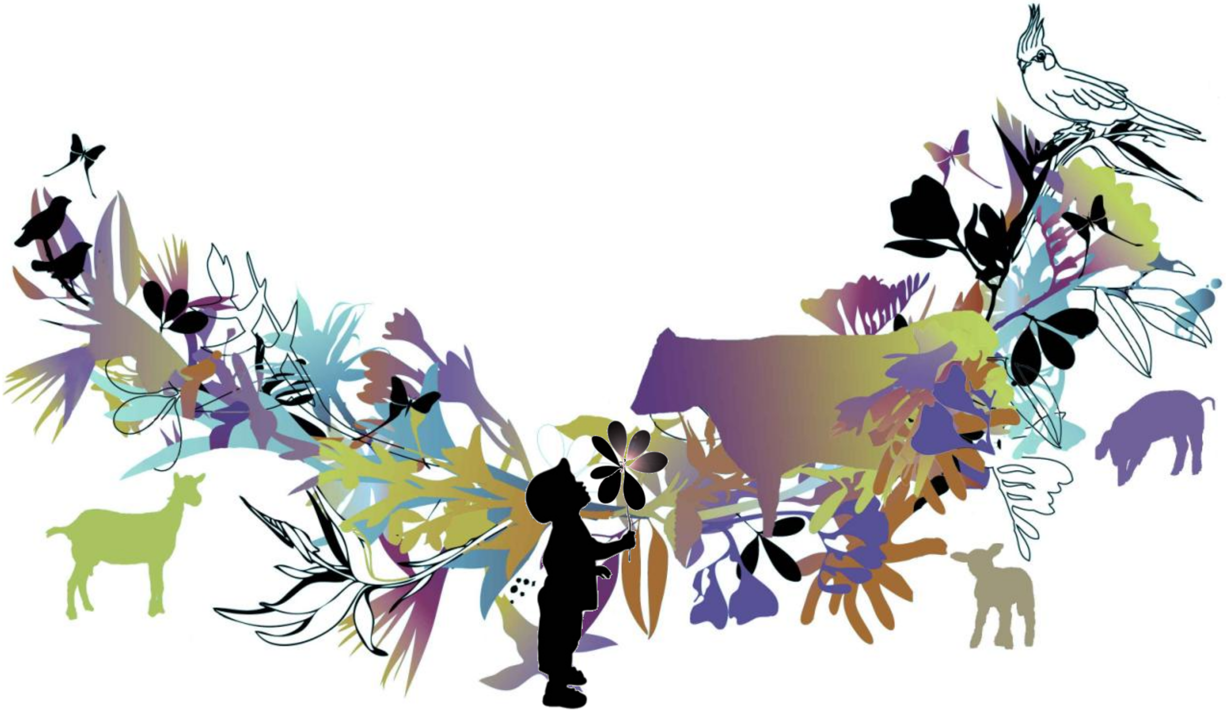
IMPUTATION ERROR RATE AND MAF



Imputation error rate in Montbéliarde

- **Have to remove poorly imputed SNPs for further analysis**
  - **May induce error in QTL detection and GS**
  - **Similar to SNP with high genotyping error rate**
- **Identification may be difficult in multi-breed context**
  - **Have to perform this analysis in all breeds**
  - **High error rate SNP in all breeds are due to mismapping**
- **Error rate in Montbéliarde drop from 0.47% to 0.37% when SNP with error rate > 5% are removed**





CONCLUSION

Conclusions

- **Imputation is efficient in most breeds of the project**
    - **When the number of HD genotype is high enough**
    - **Pay attention to poorly imputed SNP**

- **To improve the situation in other breeds :**
    - **Increase the number of HD genotypes**
    - **Through genotypes exchange in Brown Swiss and Simmental ?**

Perspectives

- **33,274 (2,562 + 30,712 imputed) HD genotypes are available**
  - **In Montbéliarde, Normande and Holstein breeds**
- **Work on this reference population with HD genotypes**
- **Within and between breeds QTL detection**
- **Within and between breeds genomic selection**

Acknowledgements

Agence Nationale de la Recherche ANR

APIS-GENE

3G UMT

INRA

UNCEIA
le réseau de la génétique animale

INSTITUT DE L'ELEVAGE

LABOGENA

M.N. Fouilloux
F. Guillaume

S. Boitard
B. Servin
M. San Cristobal
M.I. Fariello

S. Allais
A. Baur
S. Fritz

B. Salmi
D. Boichard
P. Croiseau
V. Ducrocq
F. Phocas
E. Venot
D. Laloe

Acknowledgements

Thank you for your attention

Agence Nationale de la Recherche ANR

APIS-GENE

3G UMT

INRA

UNCEIA le réseau de la génétique animale

INSTITUT DE L'ELEVAGE

LABOGENA

M.N. Fouilloux
F. Guillaume

S. Boitard
B. Servin
M. San Cristobal
M.I. Fariello

S. Allais
A. Baur
S. Fritz

B. Salmi
D. Boichard
P. Croiseau
V. Ducrocq
F. Phocas
E. Venot
D. Laloe