

Imputation of genotypes from low- to high-density genotyping platforms

Donagh Berry¹ & Francis Kearney²

¹Teagasc, ²ICBF

Motivation

- Costs ~€150 per genotype with 50,000 SNPs
- Genomics will dictate selection decisions
 - Heifers as replacements
 - Bulls as test bulls or stock bulls
- Parentage testing is a form of genotyping but in current form is not useful for genomic selection

→ reduce the cost of genomic selection

Imputation

Imputation is a method by which you try and predict the genotype of an animal at a higher density than the genotypes you actually have. This can subsequently be used to obtain genomic EBVs

Imputation

Imputation is a methhod by which you try and predict the genotype of an animmal at a highher density than the genotpes you actually have. This can subequently be used to obtain genommic EBIs

Imputation

Sire

....TCACCGCTGAG....

....CAGATAGGATT....

....??G??????A??....

....??T??????T??....

Offspring

Imputation

Sire

....TCACCGCTGAG....

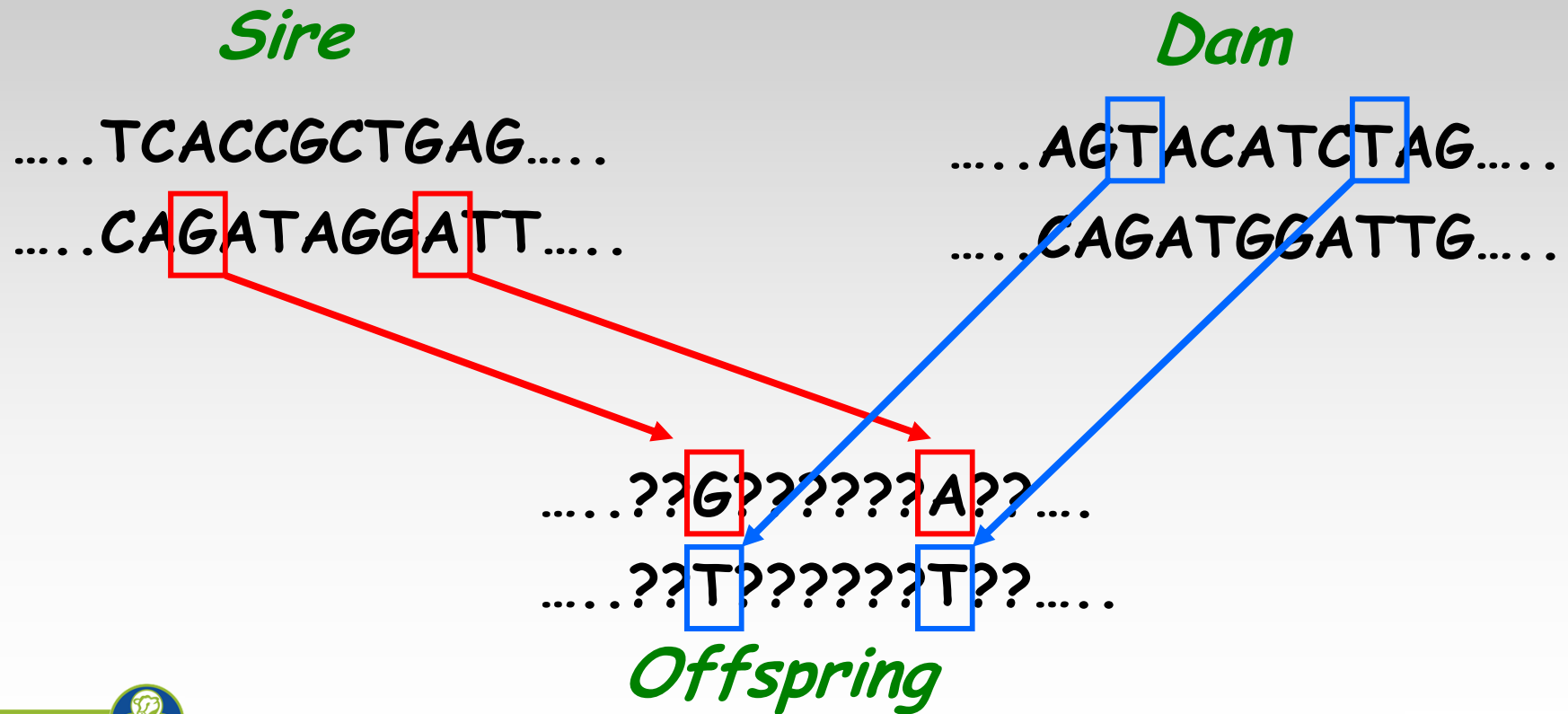
....CAGATAGGATT....

....CAGATAGGATT....

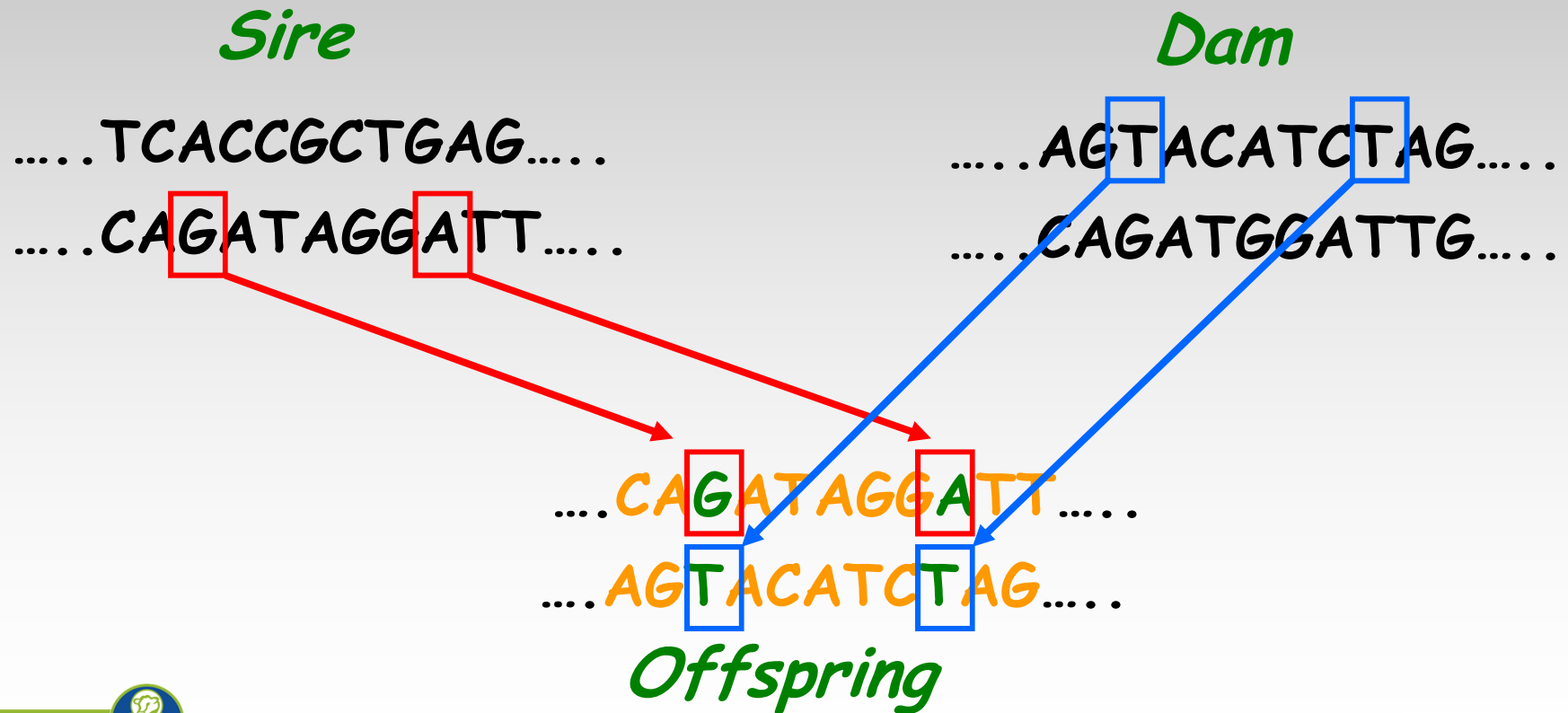
....??T?????T??....

Offspring

Imputation



Imputation



Imputation

Sire

....TCACCGCTGAG....
....CAGATAGGATT....

Dam

....AGTACATCTAG....
....CAGATGGATTG....

....CAGATAGGATT....
....AGTACATCTAG....
Offspring

Imputation

Population

MG-Sire

....AGTACATCTAG....AGTACATCTAG....
....CAGATGGATTG....CAGATGGATTG....
....AGTCGTGACTG....
.....

Sire

Dam

....TCACCGCTGAG....
....CAGATAGGATT....

....????????????????....
....????????????????....

....??G??????A??....
....??T??????T??....

Offspring

Imputation

Population

MG-Sire

....AGTACATCTAG....AGTACATCTAG....
....CAGATGGATTG....CAGATGGATTG....
....AGTCGTGACTG....
.....

Sire

Dam

....TCACCGCTGAG....
....CAGATAGGATT....

....????????????????....
....????????????????....

....CAGATAGGATT....
....AGTACATCTAG....
Offspring

Objective

To quantify the accuracy of imputing from a low- to a high-density genotype panel (54,000 SNPs) and its impact on genomic predictions

Materials and Methods

- 5,489 animals with high density genotypes
- Two groups
 - Reference group (born <2006; n=4,725)
 - Test group (born \geq 2006; n=764)
- Low density platform
 - 2,900 SNPs (2,730 after edits)
- High density platform
 - 54,001 SNPs (51,602 after edits)

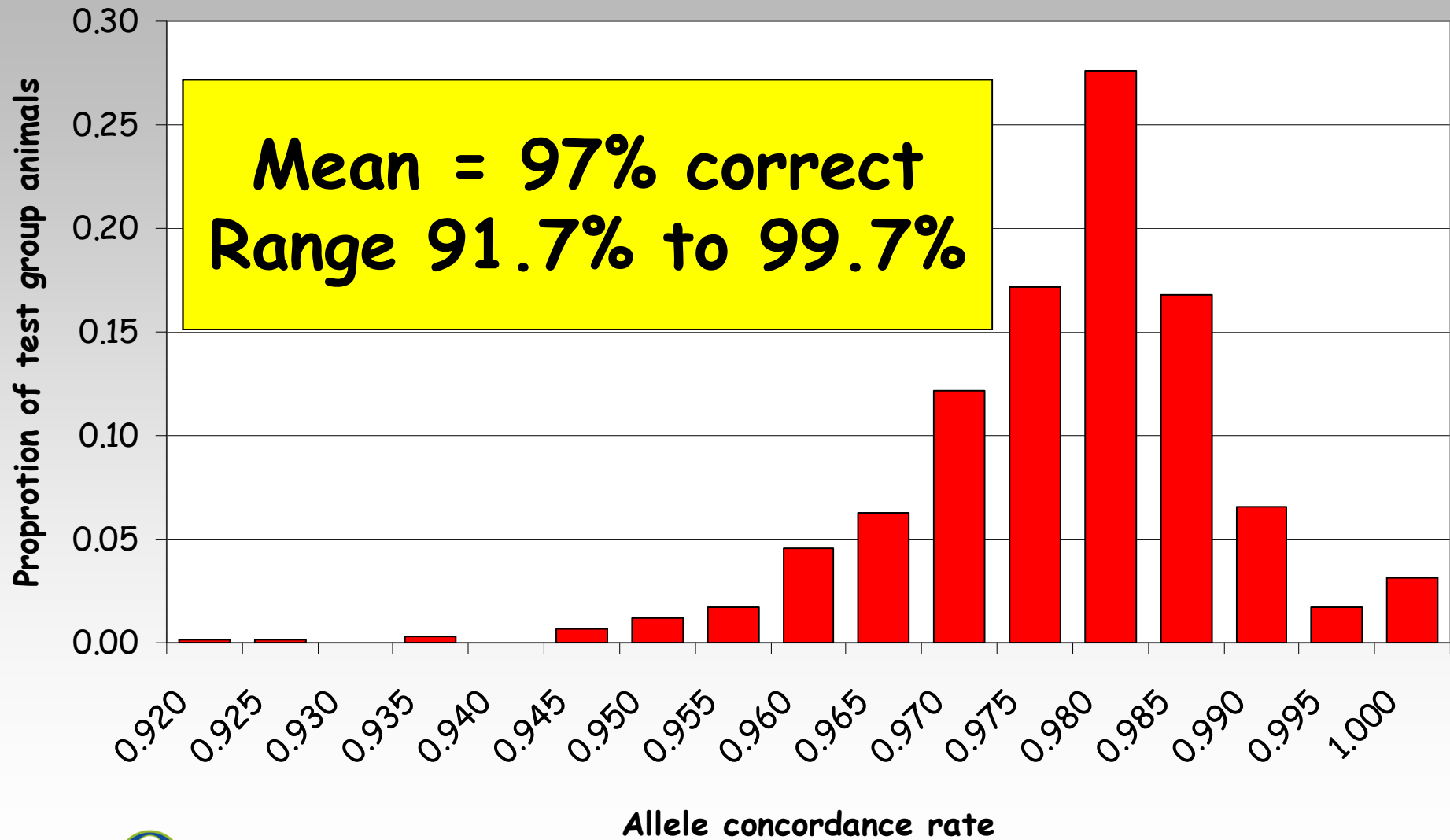
Materials and Methods

- Only low density SNPs of test group retained and remaining SNPs imputed
 - Beagle software suite
- Genomic predictions of test group animals undertaken using GBLUP
 - Predictions based on “true” or imputed genotypes
 - All traits in the EBI

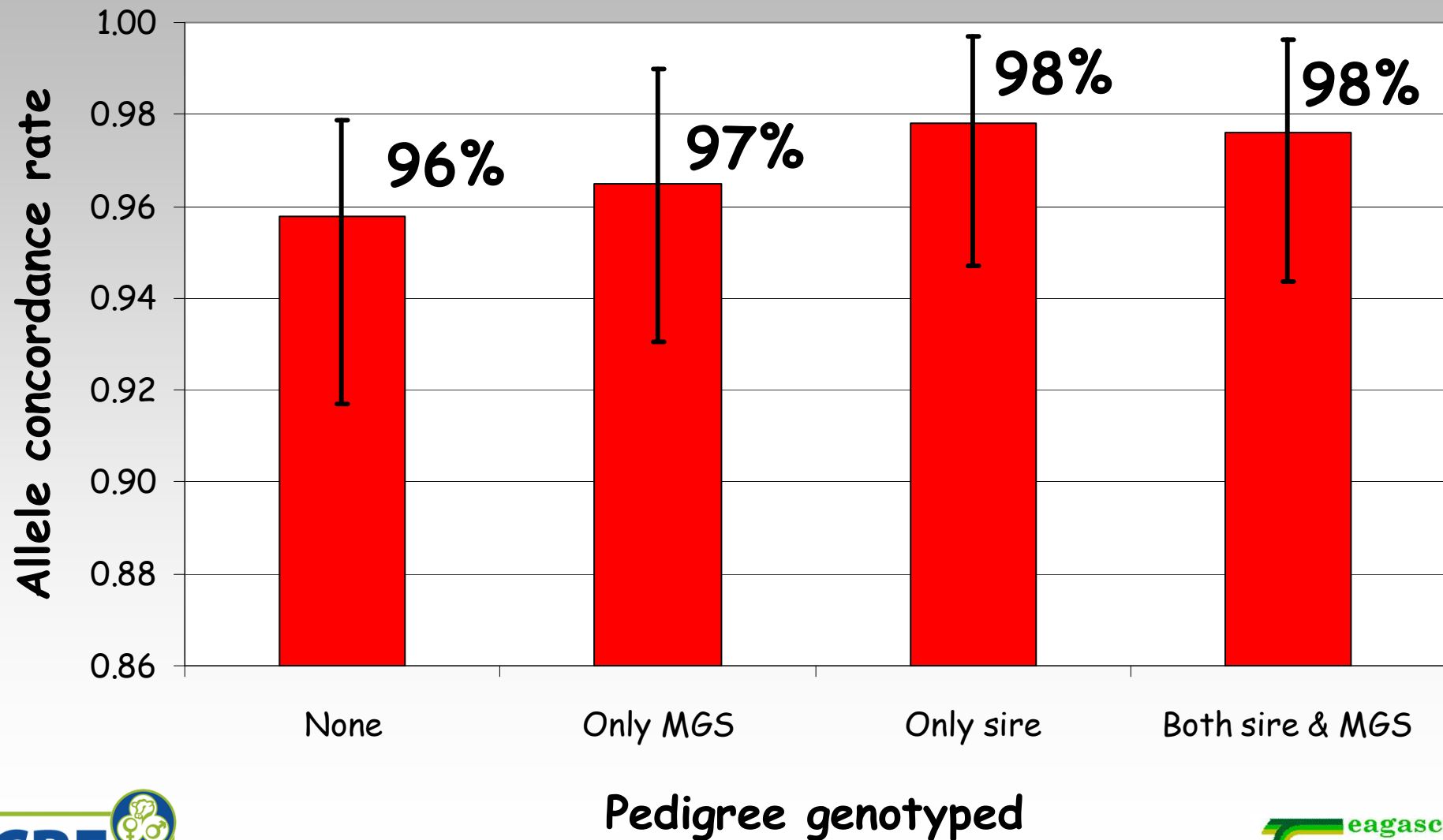
Accuracy of prediction

- **Imputation**
 - Genotype concordance rate
 - Allele concordance rate
- **Genomic predictions (comparison of direct genomic values from true or imputed genotypes)**
 - Mean bias
 - Variation in bias
 - Correlation
 - Regression

Allele concordance rate per animal



Accuracy by pedigree genotyped



Impact on genomic predictions

- Statistically significant bias of prediction (not systematic across traits) but biologically not important
- Average correlation between direct genomic values using “real” or imputed genotyped was 0.96
- Regressions of direct genomic values from “real” onto imputed genotypes close to one

Conclusions

- Accuracy of imputation is high particularly with back pedigree is genotyped
- Excellent concordance with genomic predictions using real genotypes

Implications

Framework and logistics to undertake genomic selection using the lower cost, lower-density panel (3K panel) now offered by the ICBF (€50)



This work was carried out as part of the RobustMilk project that is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708

www.robustmilk.eu