Imputation of medium density genotypes from custom low density genotype panel in sheep

D.P. Berry[1], A. O'Brien[1], S. Randles[2], K. McDermott[2], E. Wall[2], & N. McHugh[1]

[1] Teagasc, Moorepark, Ireland
[2] Sheep Ireland, Ireland

Donagh.berry@teagasc.ie

Motivation

- Aim to have multi-breed genomic evaluations in Irish sheep in 2017
- Accurate genomic predictions require a large reference population
- Average lamb value < cattle value

Need a low cost strategy for genomic selection

Imputation = "prediction of missing values"

i a a c m g h z q
a b e h k z i
y u f x b h j l

M___a_l is g_ing to g_ to _i_b__t_
ne_t y_ar in s__ing

Imputation = "prediction of missing values"

i a a c m g h z q
a b e h k z i
y u f x b h j l

John and Michael are good friends. Michael's uncle lives in Timbuktu

M___a_l is going to go to _i_b__t_
next year in spring

Imputation = "prediction of missing values"

i a a c m g h z q
a b e h k z i
y u f x b h j l

John and Michael are good friends. Michael's uncle lives in Timbuktu

Michael is going to go to Timbuktu next year in spring

Imputation = "prediction of missing values"

i a a c m g h z q
a b e h k z i
y u f x b h j l

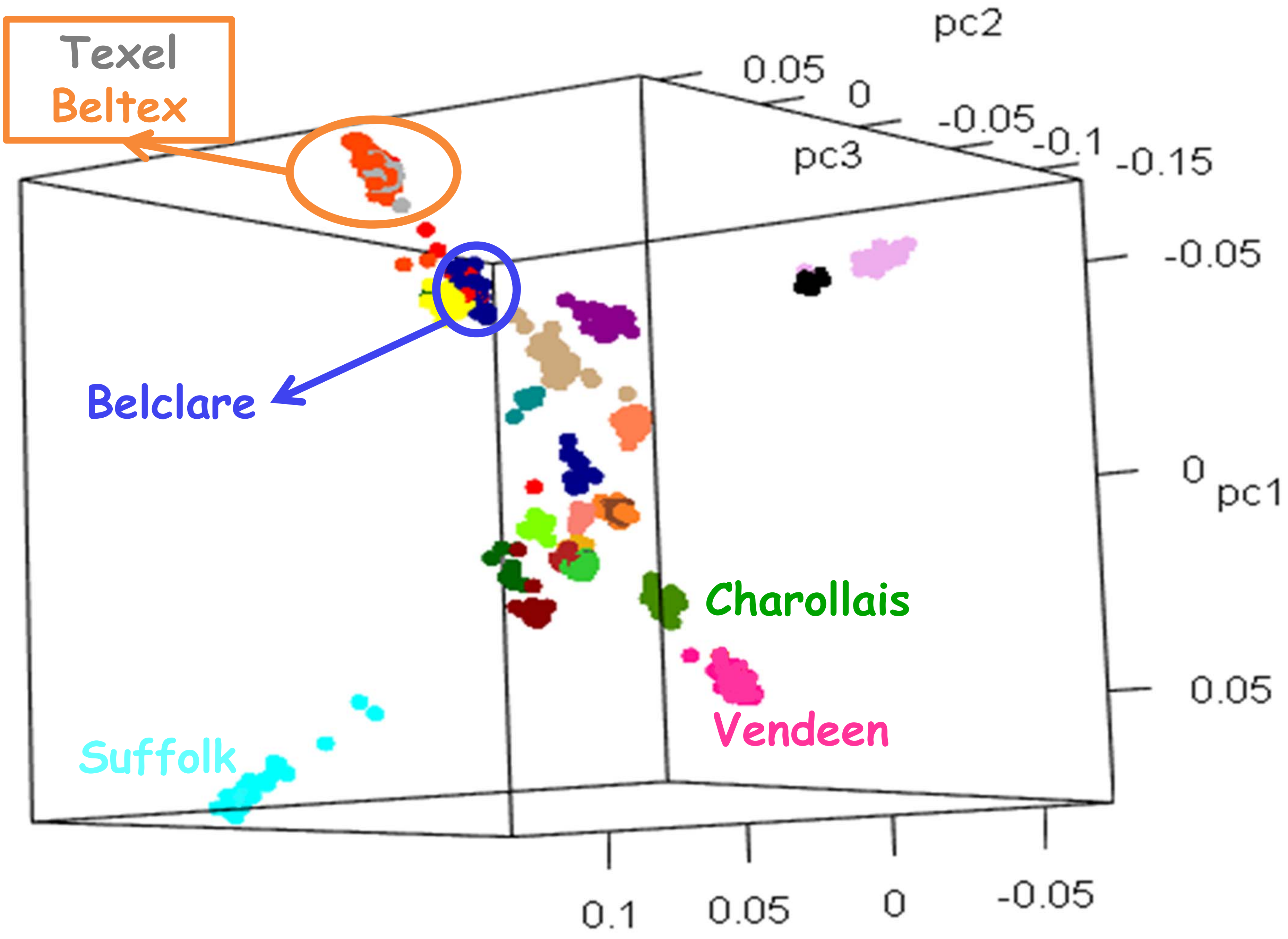
John and **Michael** are good friends. Michael's uncle lives in **Timbuktu**

M**y**chael is going to go to **Timbuktu** next year in spring

Study objective

- Current thinking is at least a 50K panel (i.e., 50,000 markers) is necessary

- Generate several thousand "50K genotypes" at a low cost
  - Many breeds
    - Texel, Suffolk, Vendeen, Charollais, Belclare


Texel
Beltex
Belclare
Charollais
Vendeen
Suffolk
pc2
0.05
0
-0.05
-0.1
-0.15
pc3
-0.05
0
pc1
0.05
0.1
0.05
0
-0.05

Study objective

- Current thinking is at least a 50K panel (i.e., 50,000 markers) is necessary

- Generate several thousand "50K genotypes" at a low cost
  - Many breeds
    - Texel, Suffolk, Vendeen, Charollais, Belclare (+Beltex)
  - Scottish friends also shared Texel genotypes

Low density panel

- 15,000 SNPs
- Equally spaced SNPs across the genome for imputation
- Greater density in putative QTL regions
- Several parentage panels
- (Putative) causal mutations
  - Scrapie, spiderlamb, GDF, Booroola, Inverdale….

Strategy

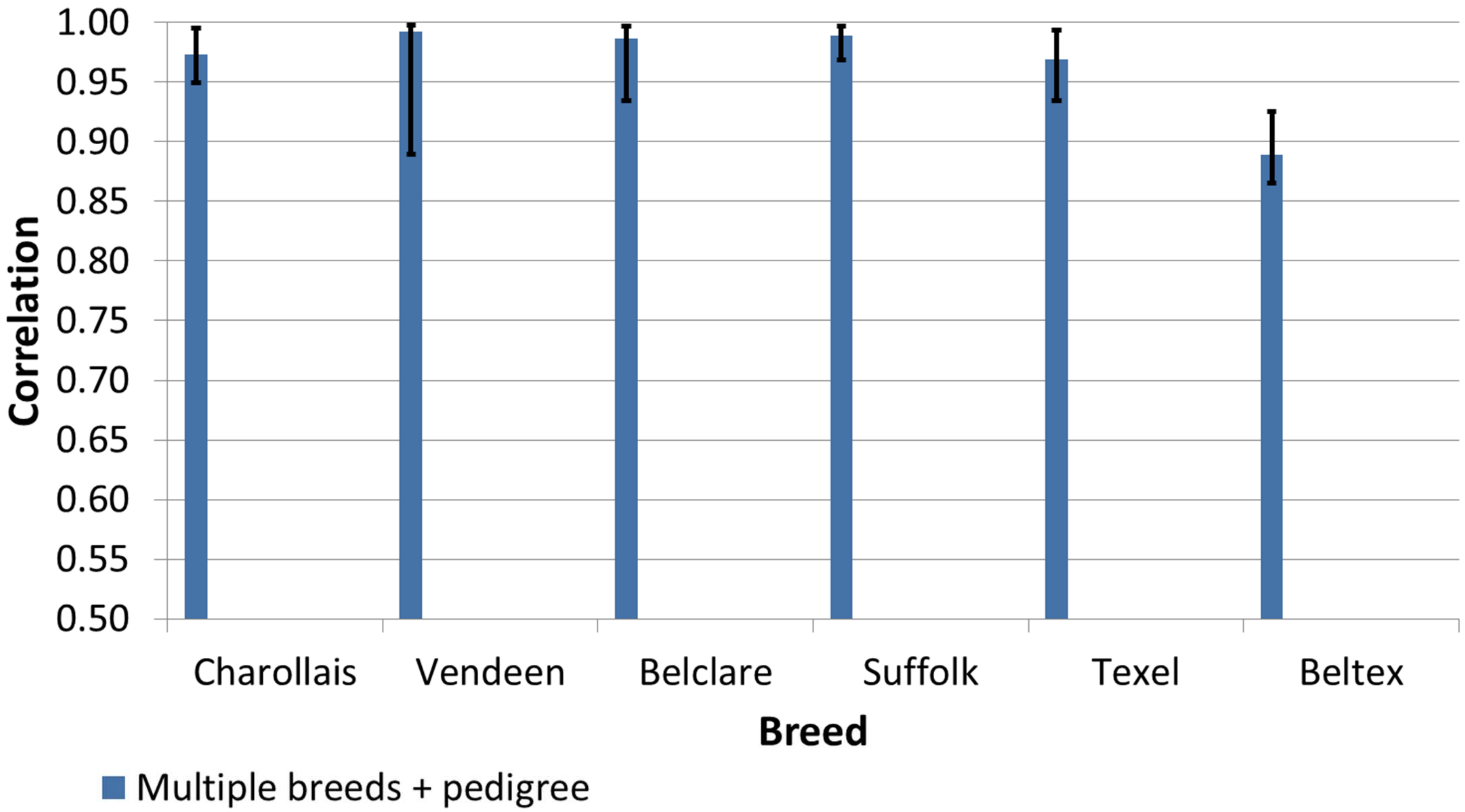
- Imputation accuracy of low density to medium density

- Reference population
  - Multi-breed Irish genotypes
  - Within-breed Irish genotypes
  - Multi-breed Irish+Scottish genotypes

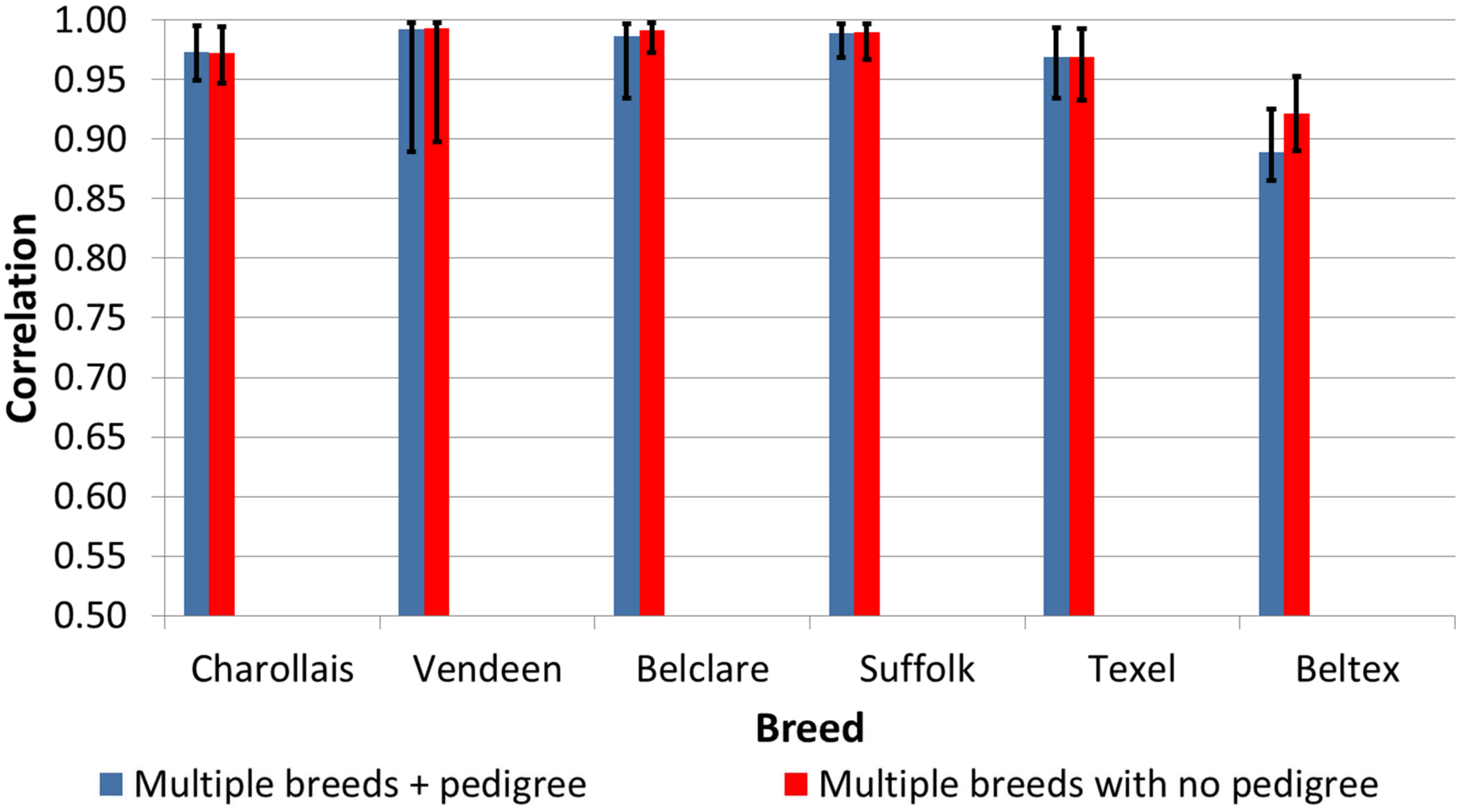
Data

Breed	Reference	Validation
Belclare	559	75
Charollais	564	75
Suffolk	684	75
Vendeen	551	75
Texel	393 + 1796	75
Beltex		64

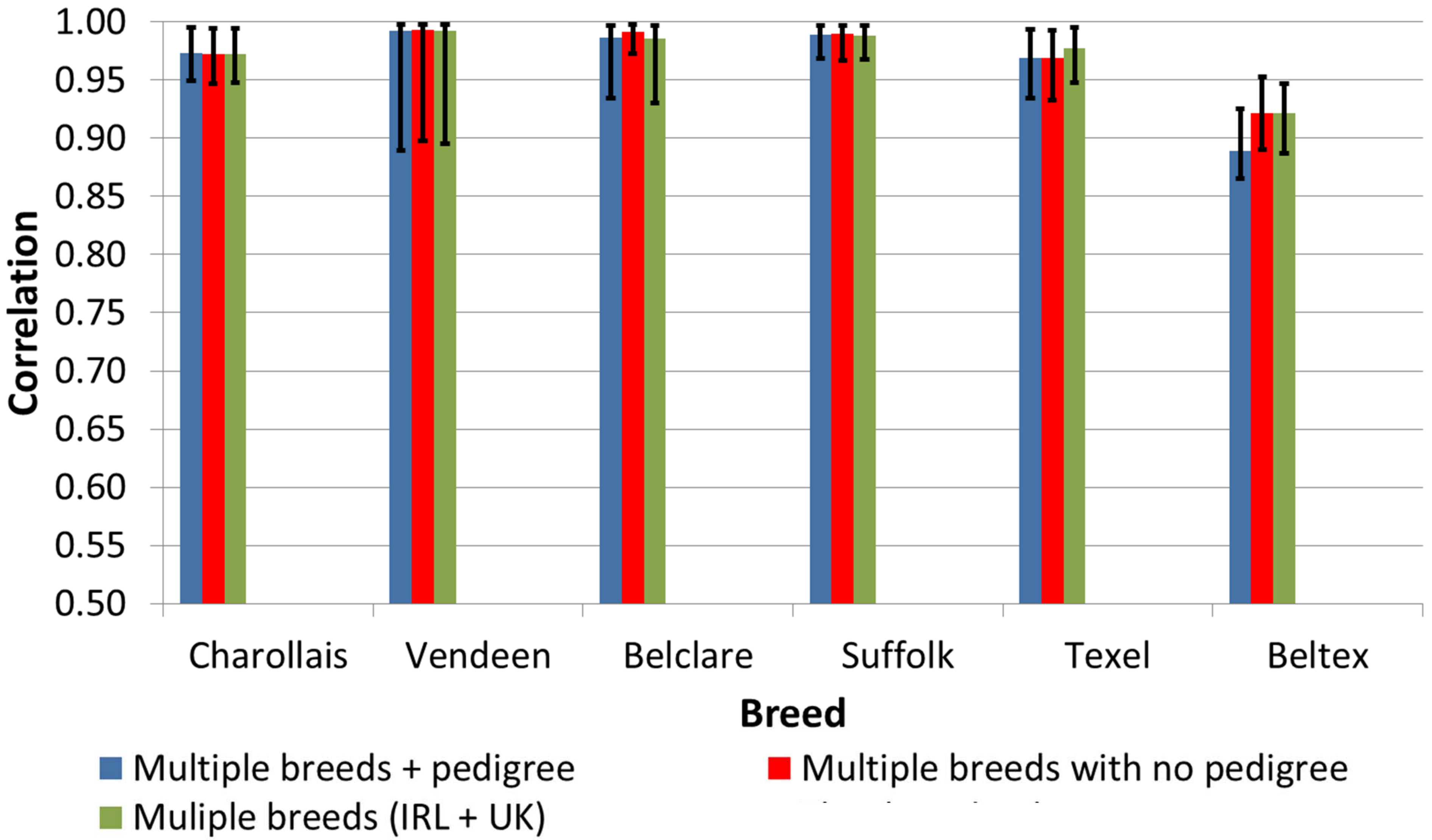
Accuracy of imputation



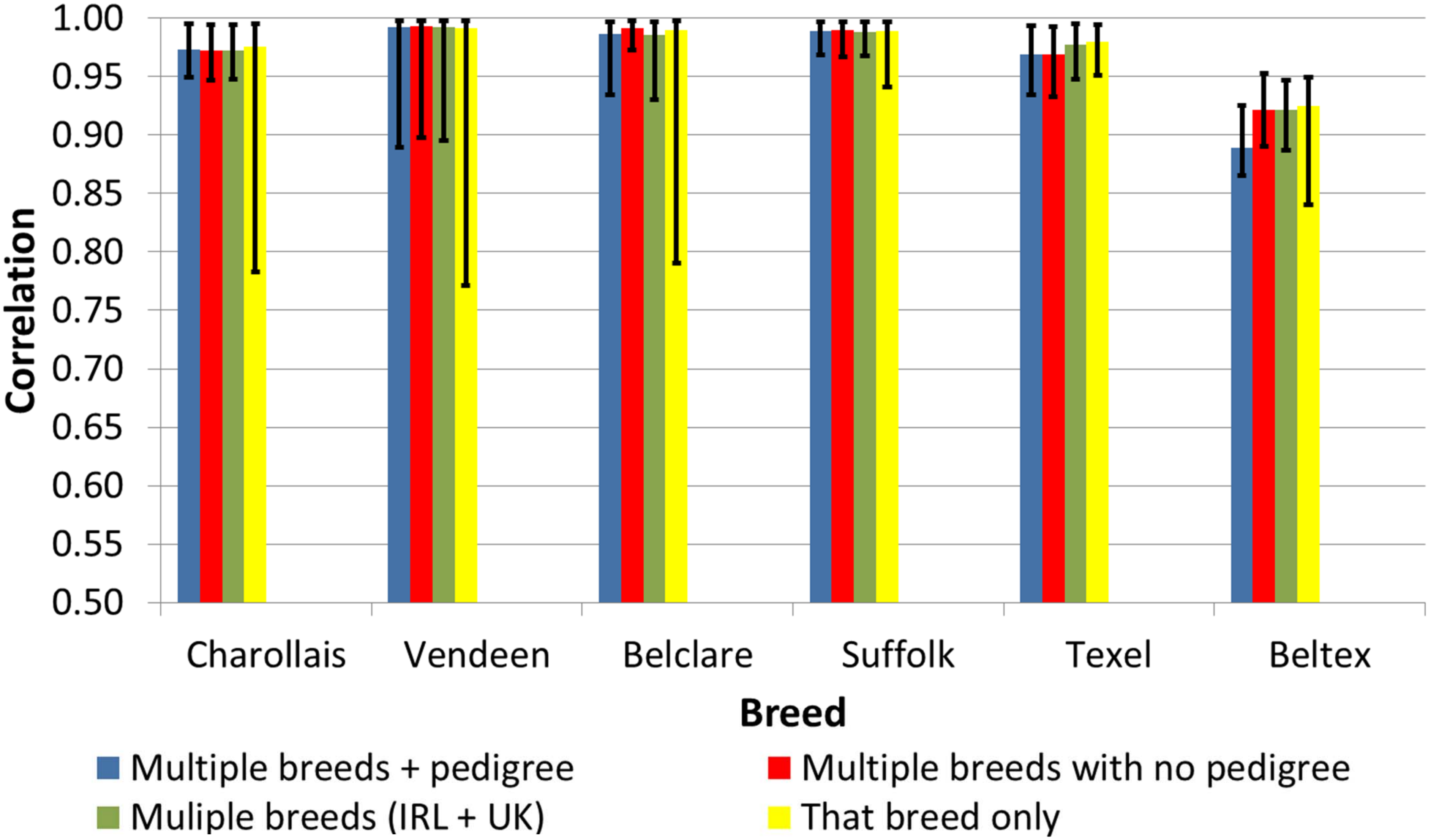
Accuracy of imputation



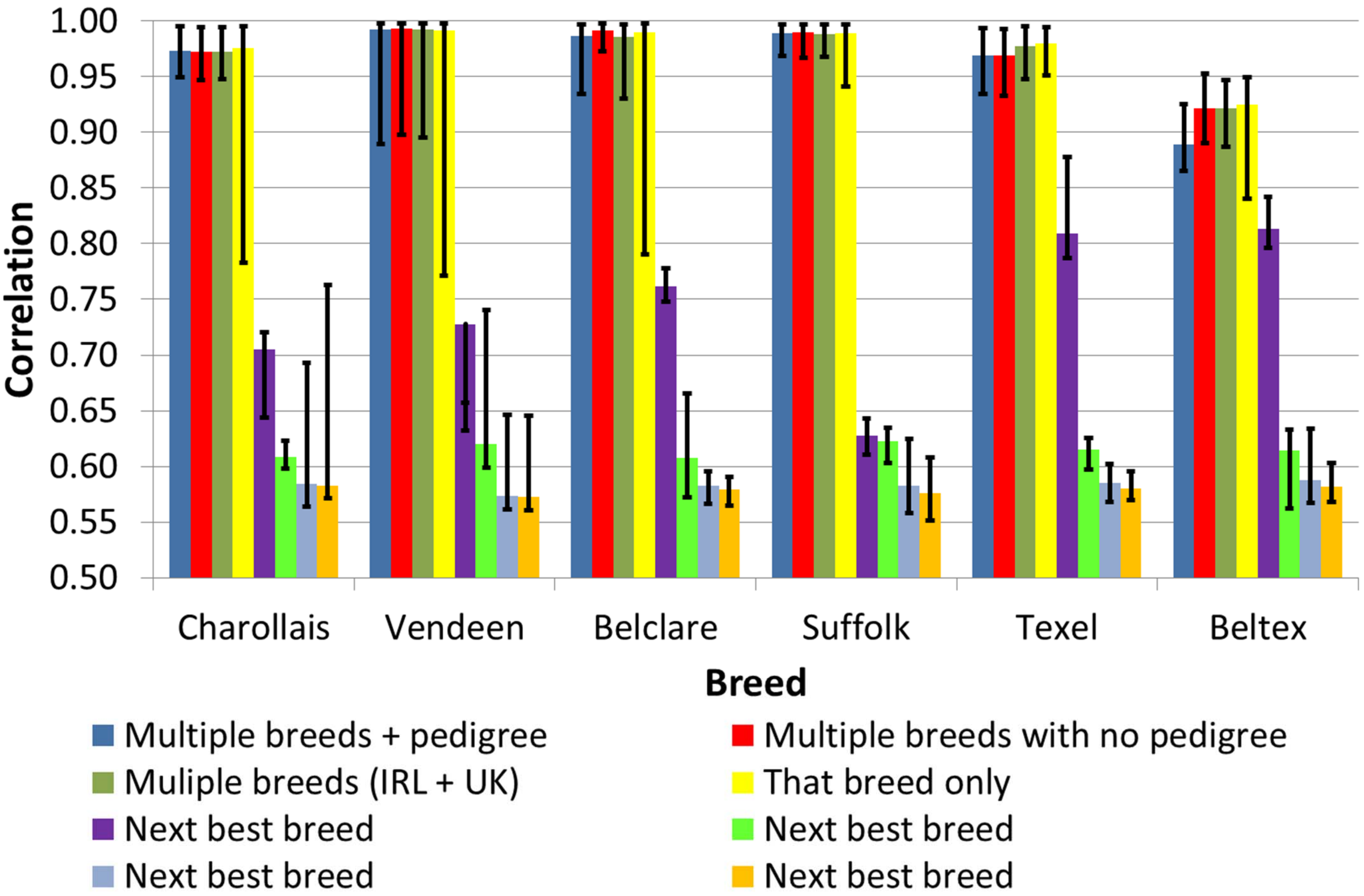
Accuracy of imputation



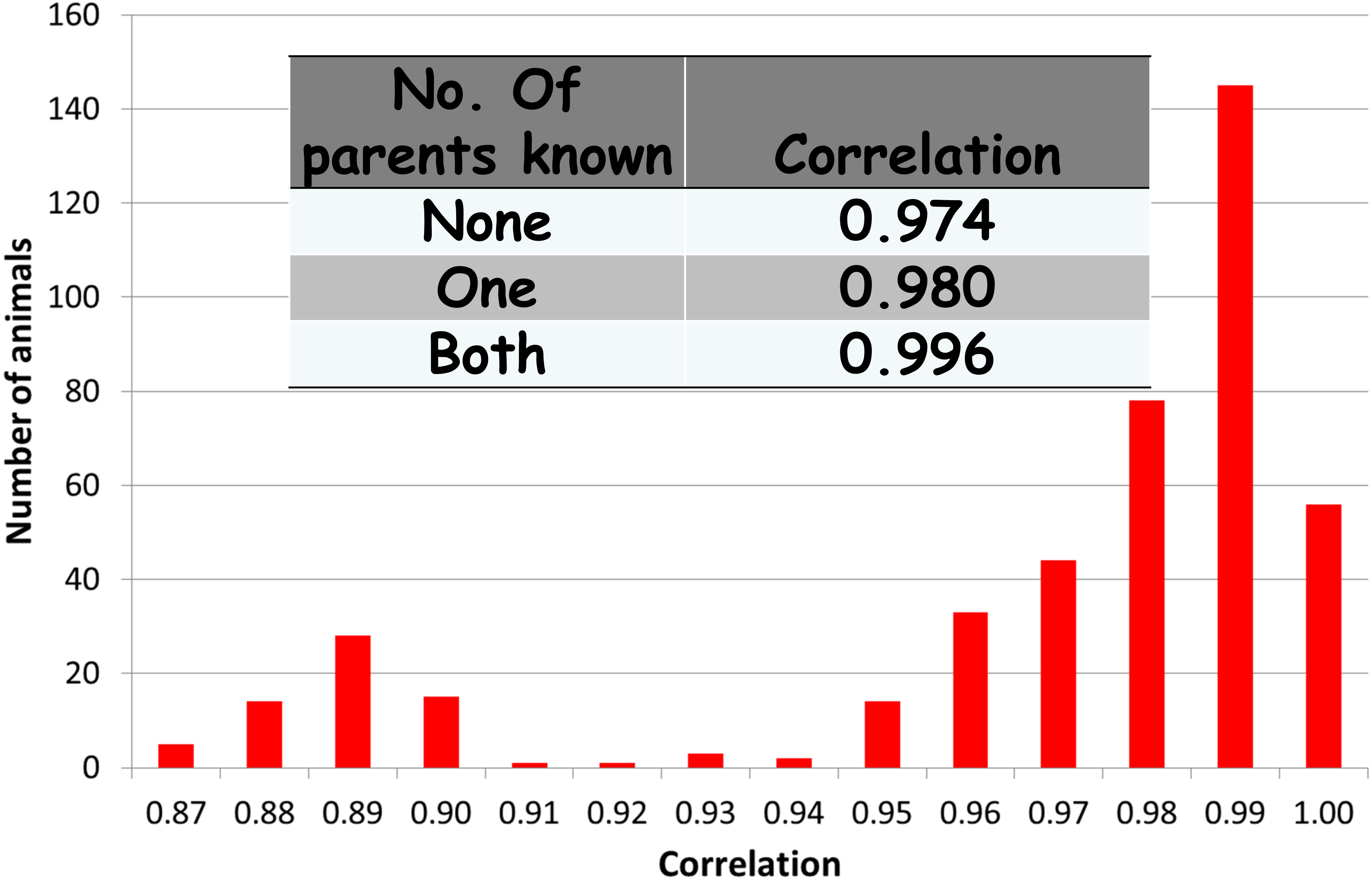
Accuracy of imputation



Accuracy of imputation



Accuracy per animal



No. Of parents known	Correlation
None	0.974
One	0.980
Both	0.996

Conclusions

- Information content on low density panel is excellent
  - Almost all SNPs segregating
- High accuracy of imputation from low to high density
  - No impact of imputing all breeds in one step
    - Probably better off to avoid the “rogues”
- Greater accuracy when parents (haplotypes) in the reference population

Acknowledgments

- SRUC & UK Texel for provision of genotype data
- International Sheep genomics Consortium (ISGC)
- Financial support from Research Stimulus Fund – OVIGEN