ARC • LNR

Excellence in Research and Development

Establishing genomic predictions in South African Holstein and Jersey cattle with small and predominantly female reference populations

K.S. Mafolo, F.W.C. Neser & M.L. Makgahlela

mafolok@arc.agric.za

Dairy Cattle Routine Genetic evaluations

- Established national routine genetic evaluations in 1987
- Participated in the Interbull's Multiple Across Country Evaluation (MACE) since 1999
  - Breeds: Holstein, Jersey & Ayrshire
- SA's AI company, Taurus established in 1978 was acquired by Evolution to form Taurus Evolution.

ARC • LNR

Excellence in Research and Development

Dairy cattle genomic evaluations?

- Integrating genomic data into genetic evaluations remain slow for several reasons:
  - High cost of setting up large reference populations, which underpins accuracy of genomic selection
  - Limited availability of local proven bulls
  - AI market-share dominated by international companies, most unwilling to share semen for genotyping or genomic data

- AI market dominated by genomic proven bulls evaluated under different environments that may not perform as well under SA environmental conditions

Step towards genomic evaluations

- Dairy Genomics Programme (DGP) was initiated in 2016 to introduce genomic selection in dairy cattle breeds, based on the expression of genes under local environmental conditions
  - Partnership between the ARC, Universities, AI companies (only Taurus Evolution and Genimex) & the dairy production industry funded by the state
  - The objectives of the DGP:
    - Setting up reference populations for the Holstein, Jersey and Ayrshire cattle
    - Measure new scarce/difficult to measure traits e.g. fertility measurements from AI service events and mastitis
    - Genomics Research (Landscape genomics, GWAS, genomic evaluations)

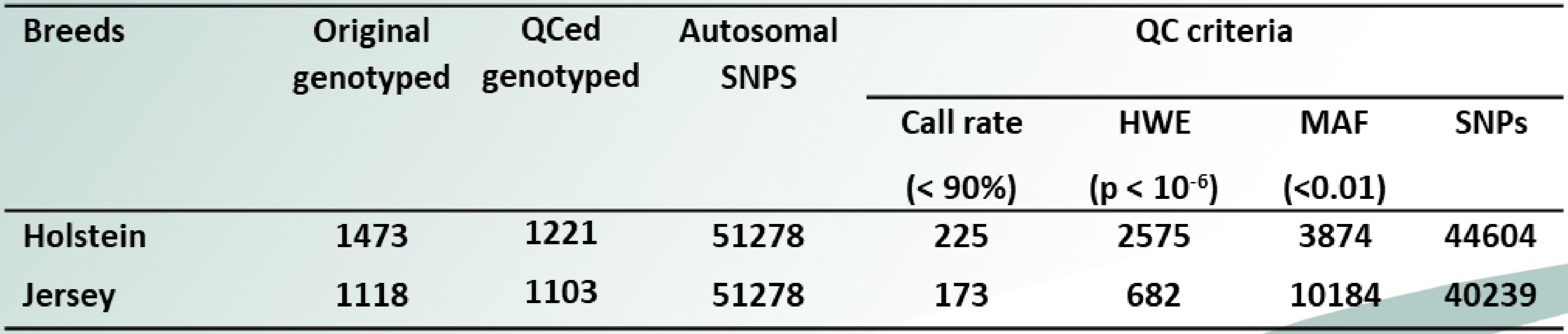
Objectives

- Investigate the composition of the SA Holstein and Jersey reference populations by assessing the genetic structure and relatedness within the genotyped animals
- Assess the accuracy of genomic evaluations for production traits.

Pedigree, phenotype data and editing

Breeds	Pedigree		305-day production data	
	Original	Edited	Original	Edited
Holstein	3 699 231	2 386 340	4 779 369	730 360
Jersey	1 712 047	999 172	3 299 495	592 491

Data were edited following the national routine evaluation procedures (https://interbull.org/ib/geforms)



Breeds	Original genotyped	QCed genotyped	Autosomal SNPS	QC criteria			
				Call rate (< 90%)	HWE $(p < 10^{-6})$	MAF (<0.01)	SNPs
Holstein	1473	1221	51278	225	2575	3874	44604
Jersey	1118	1103	51278	173	682	10184	40239

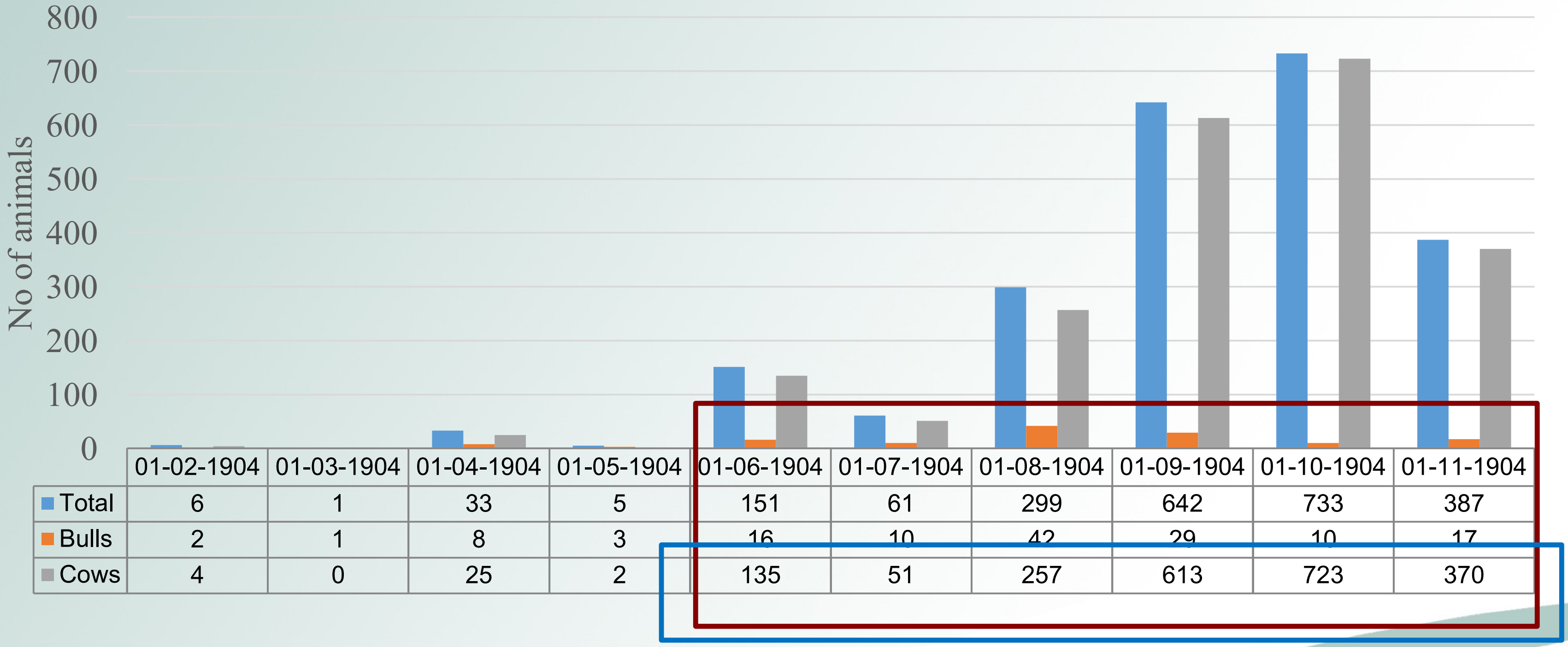
Genotyped using Illumina 50K chip v3 (Illumina Inc., San Diego, CA)
Quality control (QC); Hardy Weinberg deviation (HWE); Minor allele frequency (MAF)

Composition of the genotyped animals

		Holstein	**Jersey**
Sex	Bulls	78	60
	Cows	1 143	1 036
Pedigree	Full	1 219	1 096
	Single parent	3	7
Lactation records of genotyped cows	One Lactation	90	1
	Two Lactations	104	7
	Three Lactations	275	972
Genotyped bulls with daughters	Total	11	13
	Min no of daughters	2	3
	Max no of daughters	31	91
Dams of genotyped animals	With both genotypes and phenotypes	56	69

ARC • LNR
Excellence in Research and Development

Birth year and sex distribution of genotyped animals



	01-02-1904	01-03-1904	01-04-1904	01-05-1904	01-06-1904	01-07-1904	01-08-1904	01-09-1904	01-10-1904	01-11-1904
🟦 Total	6	1	33	5	151	61	299	642	733	387
🟧 Bulls	2	1	8	3	16	10	42	29	10	17
⬜ Cows	4	0	25	2	135	51	257	613	723	370

Most genotyped animals were young (selection candidates) and females

Definition of the reference and validation population

Code	Breed	All genotyped animals	Reference data	Validation data
CR100	Holstein	1 221	1 121	100
	Jersey	1 103	1 003	100
CR150	Holstein	1 221	1 071	150
	Jersey	1 103	953	150
CR200	Holstein	1 221	1 021	200
	Jersey	1 103	903	200
CR300	Holstein	1 221	921	300
	Jersey	1 103	803	300
CR390	Holstein	1 221	831	390
	Jersey	1 103	713	390

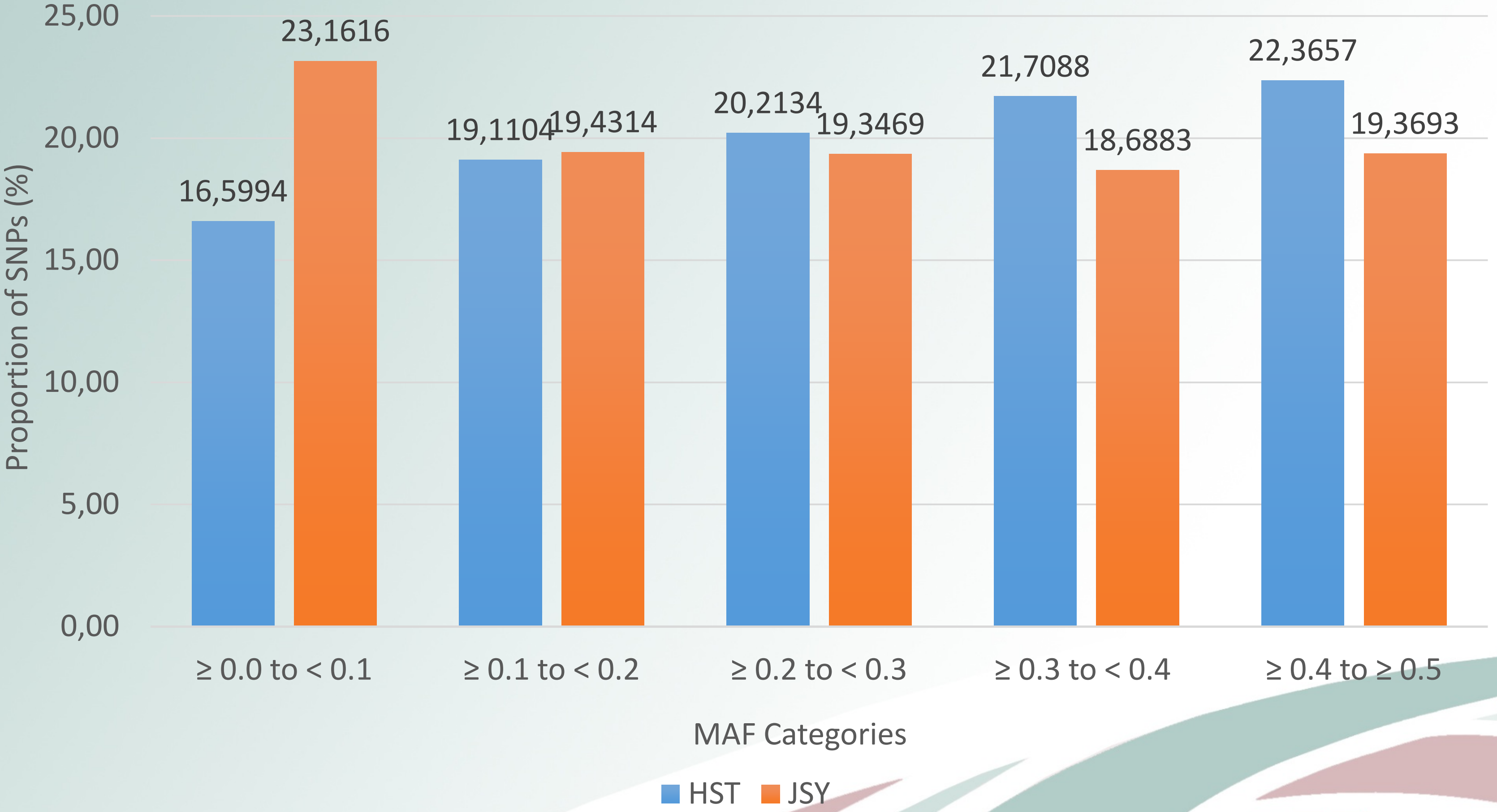
ARC • LNR
Excellence in Research and Development

Estimation of genetic and genomic evaluations

1. Pedigree-based best linear unbiased prediction (ABLUP)	**2. Genomic best linear unbiased prediction (GBLUP)**	**3. Single-step GBLUP (ssGBLUP)**
$y = Xb + Za + Wpe + e$ where **y** is the matrix of observations (305d milk yield, protein yield, fat yield, protein percentage and fat percentage) for the traits; **X**, **Z** and **W** are the known incidence matrices for fixed, random effects and permanent environmental effects, respectively. b is the vector of fixed effects; a is the vector of additive genetic effects for each animal, **pe** is the vector of permanent environmental effect and e is the vector of the residual effect.	The GBLUP model used to predict the direct genomic values (DVG) of only genotyped individuals. $y = 1\mu + Zg + e$ where **y** is the vector of phenotypes (EBVs estimated with full dataset explained below), **μ** is the overall mean, 1 is vector of 1, **g** is the vector of animal effects, assuming normal distribution of $N(0, \mathbf{G}\sigma_g^2)$, in which σ_g^2 is the variance of additive genetic effects, and **G** is the marker-based genomic relationship matrix constructed following VanRaden (2008) .	Similar to ABLUP, but the ssGBLUP the relationship matrix is replaced by the **H** matrix (Aguilar et al., 2010), in which genotypes and pedigree data are combined. Thus, the inverse of the matrix **H** is: $H^{-1} = A^{-1} +$ where $\mathbf{A}^{-1}$ is the inverse of the numerator relationship matrix (**A**) including all animals; **G** is the genomic relationship matrix ; $\mathbf{A}_{22}^{-1}$ is the inverse of the **A** matrix for only genotyped animals.

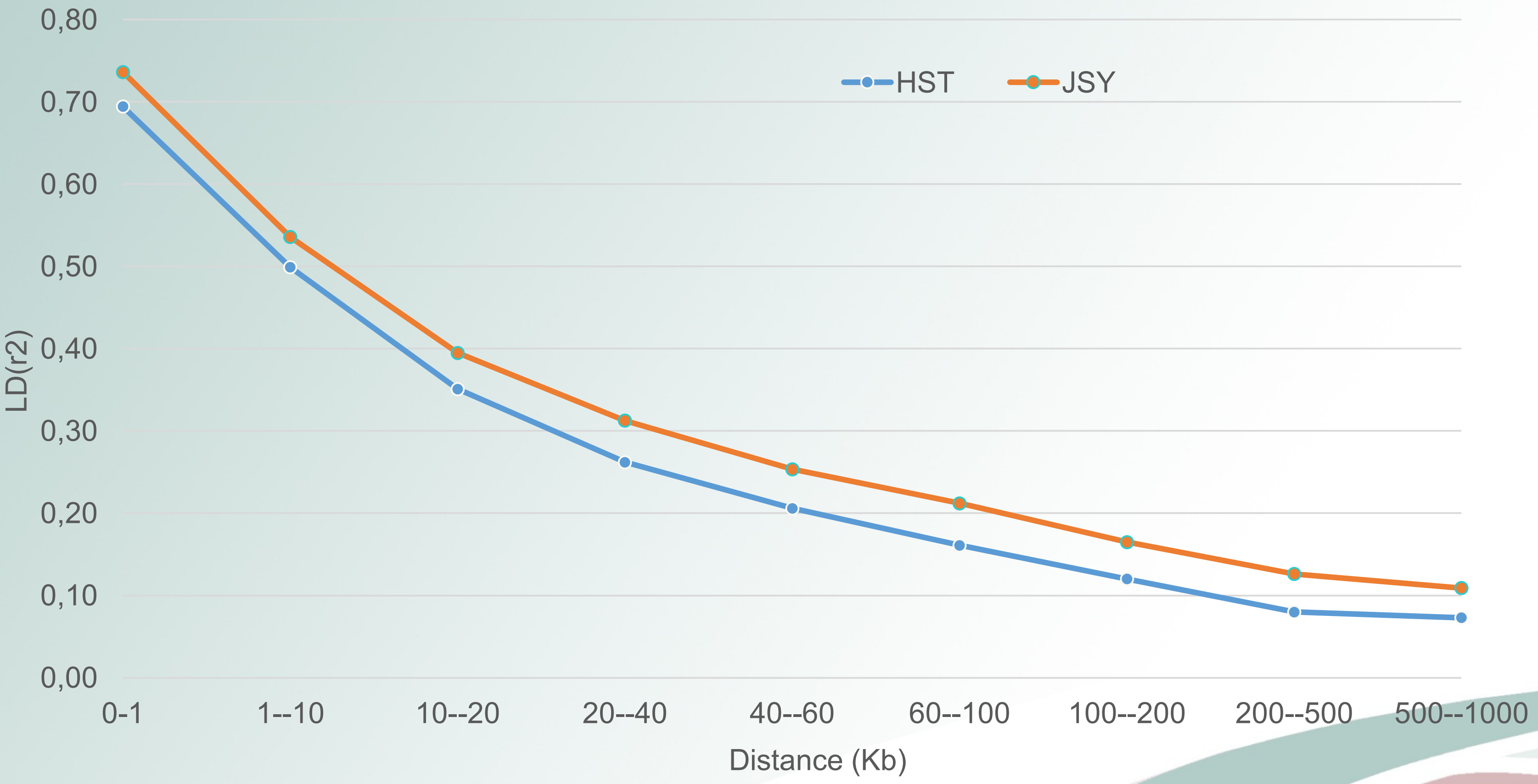
ARC • LNR
Excellence in Research and Development

Distribution of minor allele frequency



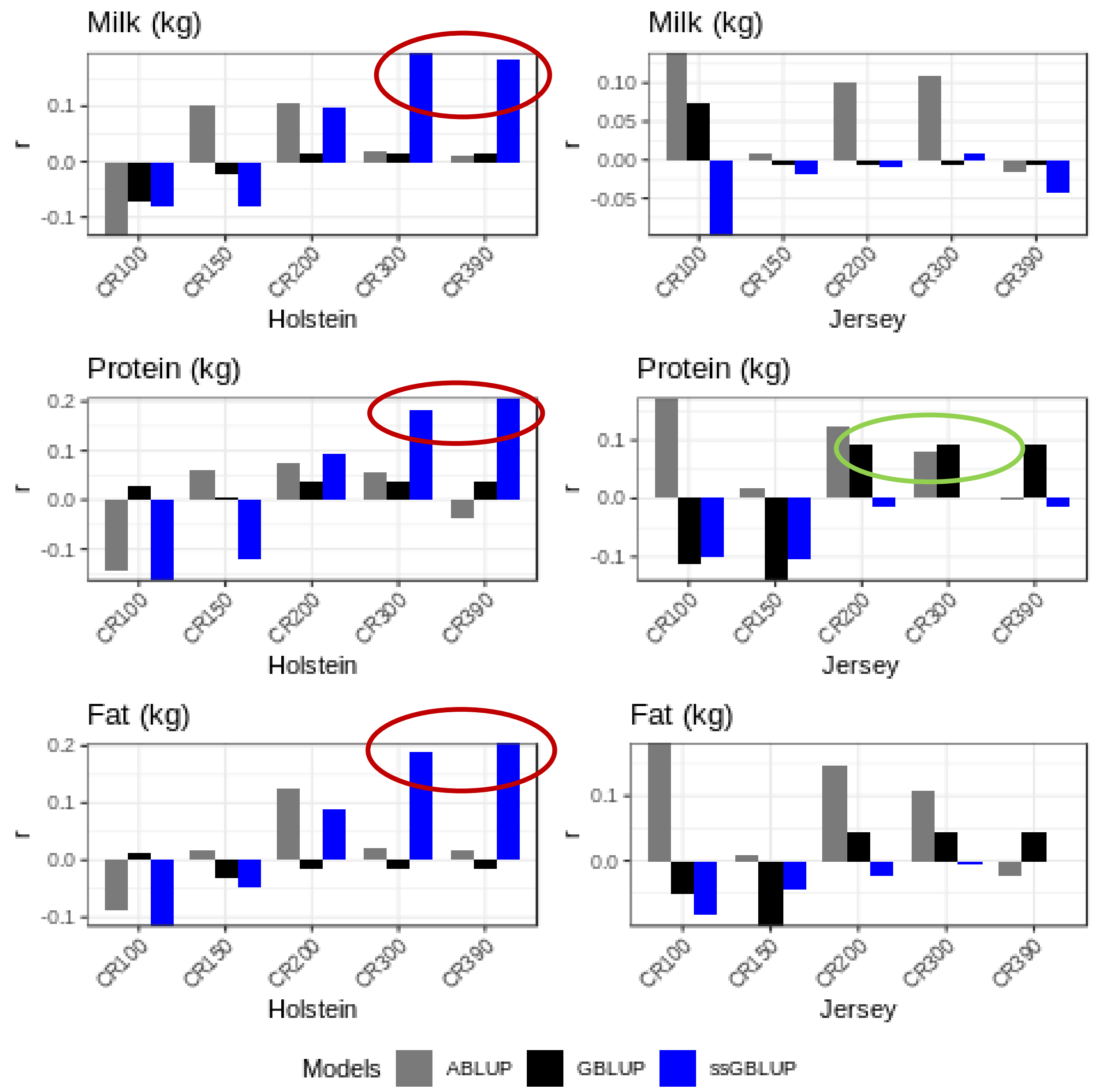
ARC • LNR
Excellence in Research and Development

Decay in linkage disequilibrium at different at different intervals



ARC • LNR
Excellence in Research and Development

Accuracy of genomic evaluations



Comparison with other developing countries

Country (Authors)	**Breed**	**Panel size**	**Traits**	**Reference population**	**Accuracy of prediction**
Brazil (Boison et al., 2017)	Gyr (Bos indicus) dairy cattle	777K (Bulls) 50K (Cows)	Milk, fat and protein yields	464 bulls 1688 cows	0.46–0.56 (only bulls reference) 0.47–0.62 (bull+cow Reference)
Kenya (Brown et al. 2016)	East Africa Crossbreds	777K	Milk yield	1038 cows	0.32–0.41 (GBLUP) and 0.28–0.35 (BayesC)
China (Ding et al., 2013)	Holstein	50K	milk yield, fat yield, protein yield	3,087 cows	0.284-0.485
South Korea (Lee et al., 2019)	Korean Holstein	50K	Milk yield, protein yield, fat yield	1,919 cows	0.32 -0.39
South Africa (Mafolo et al., 2021)	Holstein Jersey	50K	Milk yield, protein yield, fat yield	Holstein 1 143 cows 78 bulls Jersey 1 036 cows 60 bulls	Holstein -0.072 - 0.14 (GBLUP -0.16 - 0.21 (ssGBLUP) Jersey -0.078 - 0.10 (GBLUP) -0.10 - 0.10 (ssGBLUP)

Conclusion

- Accuracy of genomic evaluation were very low.
  - Possibly due to small reference population
  - Mostly young and females animals.
- Establishing the genomic infrastructure is promising towards understanding the composition of the reference population to use in genomic prediction of SA Holstein and Jersey cattle.
- The current study forms bases for more research, and improvement in the genotyping strategy.
- South Africa needs to collaborate with international countries, through which proper influential animals for more accurate evaluations could be sourced.

Acknowledgements

DGP

UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA

UFS
UV

technology innovation
AGENCY
Innovating Tomorrow Together

nicis
NATIONAL INTEGRATED
CYBERINFRASTRUCTURE SYSTEM
CHPC

NRF

Email: mafolok@arc.agric.za

ARC • LNR
Excellence in Research and Development