

## RESOURCES ASSESSMENT FOR DETERMINATION OF ANCESTRAL ALLELES IN CATTLE

J. Dorji<sup>1</sup>, L.R. Porto-Neto<sup>1</sup>, A.J. Chamberlain<sup>2</sup>, C.J. Vander-Jagt<sup>2</sup>, J. Kijas<sup>1</sup> and A. Reverter<sup>1</sup>

<sup>1</sup>CSIRO Agriculture and Food, St. Lucia, QLD, 4067 Australia

<sup>2</sup>AgriBio, Centre for AgriBioscience, Bundoora, VIC, 3083 Australia

### SUMMARY

Selection in animals whether natural or artificial leaves imprints on the genome also known as selection signatures. Such signals can pinpoint genomic regions that have undergone fixation during the selection for production, adaptation, and other domestication events. Few approaches to identifying selection signals require designation of ancestral alleles. Selection signature studies particularly at a sequence level are seldom undertaken possibly for the lack of a comprehensive list of ancestral alleles. Therefore, we reviewed the published lists of ancestral alleles in cattle and other resources of potential use in the derivation of an exhaustive list of ancestral alleles in cattle. Our results suggest the current list of ancestral alleles in cattle are few, incomplete and has low coverage on the genome. We also report on the publicly available resources particularly raw sequence reads from non-cattle *Bos* species and the 1000 Bull Genomes as readily usable resources to determine ancestral allele in cattle. Altogether, the use of genomic variants from the 1000 Bull Genomes is expected to help determine ancestral allele for about 73 million genomic positions in cattle.

### INTRODUCTION

Selection signatures are genetic imprints resulting from selection, adaptation or domestication process. The identification of selection signature is increasingly used to mine genomic regions influencing complex production and adaptation traits in cattle (Stella *et al.* 2010; Zhao *et al.* 2015; Cheruiyot *et al.* 2018). Several tools employing iHS, XP-EHH, iSAFE parameters for the detection of selection signatures require designation of ancestral alleles in the input. The ancestral alleles are alleles that persisted prior to selection and are commonly determined by comparing alleles at orthologous sites to a closely related species (Naji *et al.* 2021). However, in the absence of a list of ancestral alleles in the cattle, most if not all selection signature studies in cattle using ancestral allele dependent parameters assume the major allele as the ancestral allele in some tools (e.g., rehh package) (Gautier and Vitalis 2012). As such the major alleles are not always the ancestral allele. The minor alleles constituted more than 13 and 19% of ancestral alleles in Xiang *et al.* (2021) and Naji *et al.* (2021) respectively and such assumption can have significant influence the inferences.

In this study, we reviewed published lists of ancestral alleles and assessed genomic resources of cattle and out-species with potential for the determination of ancestral alleles. We examined the coherence of genomic position among the published lists of ancestral alleles and drew insight on the population structure of previously unused out-species.

### MATERIALS AND METHODS

We reviewed the published list of ancestral alleles in cattle for their coverage or the number of sites, associated reference genomes and the closely related species used for its determination. Wherever comparable, the number of sites in common and concordance of ancestral alleles were estimated. In terms of resource for determination of ancestral alleles in cattle, we queried the raw sequences of non-cattle *Bos* species in the NCBI-SRA. Further, we use raw sequence reads of *Bos* species that were not previously used in the determination of ancestral alleles to draw insights on the population structure in relation to cattle. The sequence reads were processed following the

pipelines used for processing the 1000 Bull Genome. Further, we explored the coincidence of variant position in the 1000 Bull Genomes with previous studies and estimated its potential contribution to the existing ancestral allele list.

## RESULTS AND DISCUSSION

**Ancestral alleles.** Up to now, at least four studies have determined the lists of ancestral alleles in cattle (Table 1). They were determined by comparing allele of cattle species to non-cattle *Bos* species and other non-Bos species. The earlier two lists of ancestral alleles were based on the older bovine reference genome (UMD3.1) and Bovine SNP panels. The third and fourth studies were based on the reference genome, ARS-UCD1.2 and independently determined ancestral allele for 32 and 40 million positions corresponding to about 25% of the total variants detected in cattle to date.

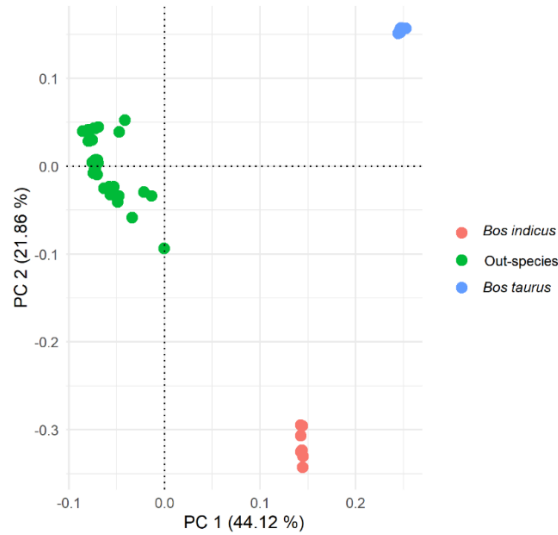
**Table 1. Summary of ancestral alleles in cattle published**

Coverage	Reference genome	Cattle spp <sup>†</sup>	Non-cattle <i>Bos</i> spp	Out-group spp	No. of AA* determined	References
BovineSNP50	UMD3.1	Taurus Indicine Composite	Gaur Banteng Yak	Bison Low-land anoa Cape buffalo	50.1 thousand	Matukumalli <i>et al.</i> 2009
BovineSNPs (19.5 million variants)	UMD3.1	Taurine	Yak	Sheep Water buffalo	14.4 million	Rocha <i>et al.</i> 2014
Whole genome sequence	ARS- UCD1.2	Taurine Indicine	Yak Banteng Gayal Gaur Auroch	Bison	32.4 million	Naji <i>et al.</i> 2021
Whole genome sequence	ARS- UCD1.2	Taurine Indicine Composite	Yak	Sheep, camel	39.9 million	Xiang <i>et al.</i> 2021

\*AA: Ancestral allele

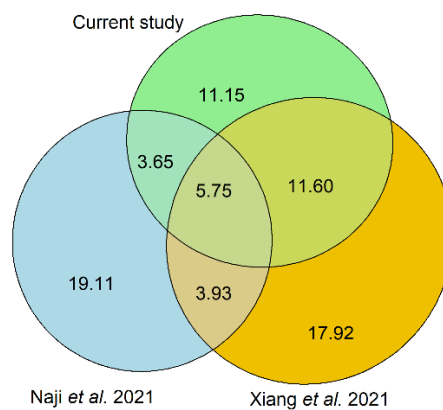
Between the two recent lists of ancestral alleles, about 9 million positions were in common. The coincidence of the ancestral alleles in the common positions were very high (99.8%). Altogether, these two lists presented ancestral alleles for about 60 million positions.

Query of raw sequence reads of non-cattle *Bos* species showed *Bos mutus* (N=4) and *Bos sauveli* (N=2) have not been previously used in the determination of ancestral allele in cattle. The inclusion of these species is likely to improve the reliability of some of the current ambiguous and low probability sites in the lists. The population structure of cattle and out-species (Figure 1) showed less prominent segregation among the out-species group compared to cattle. This is expected because the variant positions in out-species were ortholog of cattle which not necessarily segregated in the out-species. The PC1 largely separated out cattle and out-species groups while PC2 segregated *Bos indicus* and *Bos taurus*.



**Figure 1. PC plot (PC1 and PC2) of cattle (*Bos indicus* and *Bos Taurus*) and out-species**

**1000 Bull Genome.** It is a massive genomic resource collating variants for genomic imputation and genome wide association studies in cattle. The dataset has ~32 million high confidence (i.e., PASS) biallelic variant positions for cattle. It also provided genomic variants for out-species (five non-cattle *Bos* species including bison). This is a readily usable resource for the determination of ancestral allele by comparing alleles in cattle with alleles in orthologous positions in the out-species. The coherence of this genomic positions with two previous studies combined were more than 65% (Figure 2). This dataset would add another 11 million genomic positions for ancestral alleles to the existing list to reach 73 million. Further, considering the next best confidence category of variants (i.e., Tranche90to99) which is about 40 million positions can substantially increase the positions of ancestral alleles in cattle up to 100 million.



**Figure 2. Positional coherence between the two studies and variant positions in the 1000 Bull Genomes (in million)**

## **CONCLUSION**

Despite limited studies investigating on the ancestral alleles in cattle, there is a high proportion of positions in common between the studies to investigate the coherence of ancestral alleles. Further, the use of available genomic resources is expected to significantly improve the coverage of ancestral allele on the cattle genome and to enhance ancestral allele-based detection of selection signature studies in cattle.

## **ACKNOWLEDGEMENTS**

Authors sincerely thank Agriculture Victoria, Department of Energy, Environment and Climate Action (DEECA), Victoria for sharing the semi-processed sequences of out-species. We are also thankful to the 1000 Bull Genome Consortium for the meta-data and summary statistics of the 1000 Bull Genomes.

## **REFERENCES**

- Cheruiyot E.K., Bett R.C., Amimo J.O., Zhang Y., Mrode R., and Mujibi F.D.N. (2018) *Front. Genet.* **9**: 607.
- Gautier M. and Vitalis R. (2012) *Bioinformatics.* **28**: 1176.
- Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F., Heaton M.P., O'Connell J., Moore S.S., Smith T.P.L., Sonstegard T.S. and van Tassell C.P. (2009) *PLOS ONE.* **4**: e5350
- Naji M.M., Utsunomiya Y.T., Sölkner J., Rosen B.D. and Mészáros G. (2021) *BMC Genomics.* **22**: 108.
- Rocha D., Billerey C., Samson F., Boichard D. and Boussaha M. (2014) *J. Anim. Breed. Genet.* **131**: 483.
- Stella A., Ajmone-Marsan P., Lazzari B. and Boettcher P. (2010) *Genetics.* **185**: 1451.
- Xiang R., Breen E.J., Bolormaa S., Vander-Jagt C.J., Chamberlain A.J., Macleod I.M. and Goddard M.E. (2021) *Comm. Biol.* **4**: 1353.
- Zhao F., McParland S., Kearney F., Du L. and Berry D.P. (2015) *Genet. Sel. Evol.* **47**: 49.