Contributed paper

CIRCULAR GENOMIC PERMUTATIONS CAN LIMIT THE CONFOUNDING EFFECTS OF THE REFERENCE POPULATION IN THE ANALYSES OF SELECTION SIGNATURES

I.A.S. Randhawa[1,2], M.S. Khatkar[1], P.C. Thomson[1] and H.W. Raadsma[1]

[1] The University of Sydney, Sydney School of Veterinary Science, Camden, NSW, Australia
[2] The University of Queensland, School of Veterinary Sciences, Gatton, QLD, Australia

SUMMARY

Analyses of selection signature are extensively used to detect chromosomal regions underlying phenotypic diversity which have been subjected to selective pressure. However, breeds with common origin, recent divergence, or similar production types are often confounded because the candidate and reference populations, compared in such analyses, exhibit similar patterns in genomic data. This study has applied a circular genome permutation method to generate a reference population to investigate selection signatures in Angus cattle (n=29) by using 1.6 million SNPs and applying the composite selection signals (CSS) method. Significant CSS were compared in two sets of analyses based on different reference populations, i.e., CSS-1: using circular genome permutation to form the breed neutral reference population (n=29) for Angus, and CSS-2: using five beef breeds as a reference population (n=36). Notably, several genomic regions were detected using CSS-1 (e.g., on chromosome 14, 16, 21) in Angus underlying commonly known genes of major effects on beef traits which were not detected by CSS-2 because of the confounding genetic background of Angus with the reference beef breeds. The results highlight the importance of selecting an appropriate reference population to circumvent the confounding breed effects.

INTRODUCTION

In livestock species, genomic data of various breeds are compared for genotypic and haplotypic distributions against each other, individually or as groups, to characterize the historic selective pressures for breed specific traits of production, health and adaptation. These signatures of selection can be used to discover genetic variants and genes to understand the biological control of agricultural and health traits (Kemper and Goddard 2012). Genomic investigations are frequently resource-intensive; however, detection of selection signatures can provide insights into the genetic architecture underlying breed-specific traits in a relatively cost-effective manner (Gibbs *et al.* 2009). Ubiquitous cattle breeds, such as Angus have been selectively improved for economic traits by increasing the frequency of beneficial alleles throughout the genome.

A review of recently published signatures of selection showed that a few regions are commonly found in multiple breeds, suggesting genomic hotspots underlying genes of major effect, e.g., *PLAG1* on bovine autosome 14 (BTA14) (Randhawa *et al.* 2016). It was noted that the locations of selection signatures generally varied in those studies due to differences in sample size, SNP density and reference population. Rapidly lowering costs have allowed a larger number of samples to be assayed for high-throughput genotyping and genome-wide sequences. However, selection of an appropriate reference panel and thus avoiding confounding effects of the reference population remains a challenge. Inclusion of related breeds in the reference population can mask the detection of common selection signatures. This study has applied a new approach of circular chromosomal permutations (Cabrera *et al.* 2012) to generate a breed neutral reference population though permutation of the genome of the same breed used in the comparison and thus is likely to be free from breed bias. This approach was used to detect selection signatures in Angus and the results were compared with conventional breed-vs-breed approach.

MATERIALS AND METHODS

Briefly, the circular genomic permutation approach considered SNPs along each chromosome as a circular fragment, which was then rotated for each sample (animal) to pick a starting location randomly. The approach shuffles the chromosomal fragments while keeping intra-sample haplotypic and linkage structures. All permuted samples were assembled and considered as a reference population for the same breed used in the comparison, expecting genome-wide neutral and uniform genetic diversities. The composite selection signal (CSS) method (Randhawa *et al.* 2014) was used to detect across-breed selection signatures in two data sets, i.e., CSS-1: Angus vs Permuted reference genome (using the 29 Angus samples), and CSS-2: Angus vs 5 beef breeds as reference. A total of 65 beef cattle samples (Table 1) and ultra-high density genotypic data (1,583,288 SNPs) were used.

Table 1. Cattle breeds, their geographic origin, country of sampling and DNA samples

Breeds	Type	Geographic origin	Country of sampling	Samples
Angus	Beef	Scotland	USA, New Zealand, Australia	29
Brahman	Beef	India	India, Brazil, USA, Australia	8
Hanwoo	Beef	Korea	South Korea	11
Murray Grey	Beef	Australia	Australia	1
Simmental	Beef	Switzerland	USA	6
Wagyu	Beef	Japan	USA	10
Total	**-**	**-**	**-**	**65**

RESULTS AND DISCUSSION

Genome-wide analyses detected a number of expected genomic regions in Angus by CSS-1 (within breed), while CSS-2 (between breeds) missed most of the regions, as shown by three example chromosomes (BTA14, BTA16 and BTA21) in Figures 1, 2 and 3.

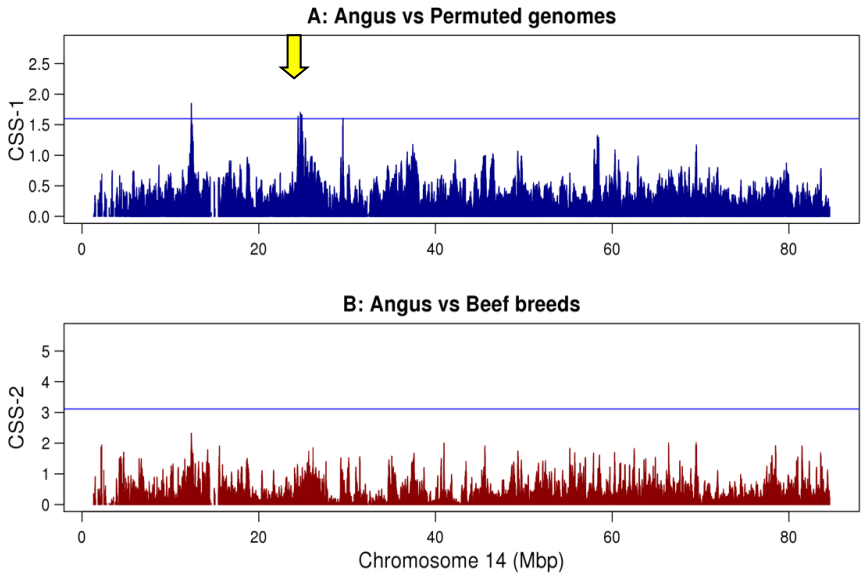

Figure 1. Composite selection signals on chromosome 14 in Angus cattle, computed by using two different reference populations; A) Permuted genomes, B) Beef breeds. Blue lines at genome-wide top 0.1%. Arrow at the top shows position of the expected signatures of selection in Angus

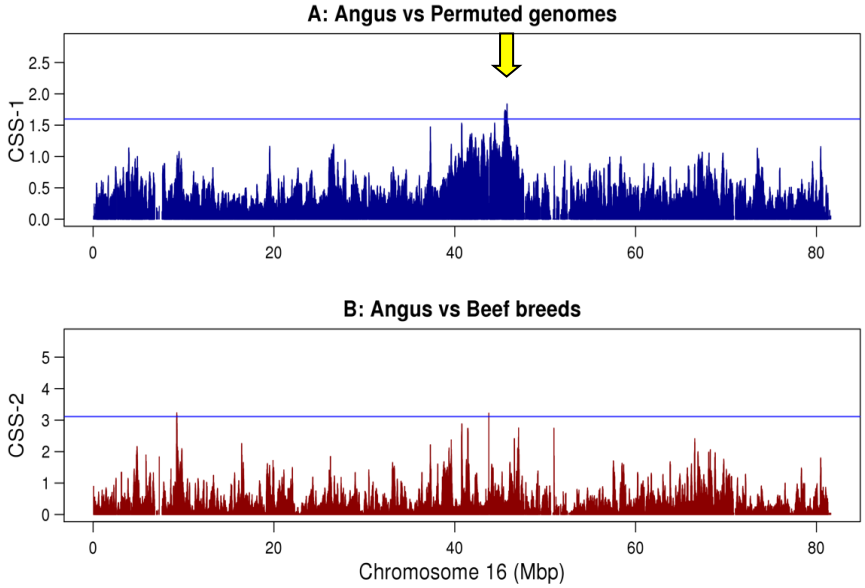

Figure 2. Composite selection signals on chromosome 16 in Angus cattle, computed by using two different reference populations; A) Permuted genomes, B) Beef breeds. Blue lines at genome-wide top 0.1%. Arrow at the top shows position of the expected signatures of selection in Angus

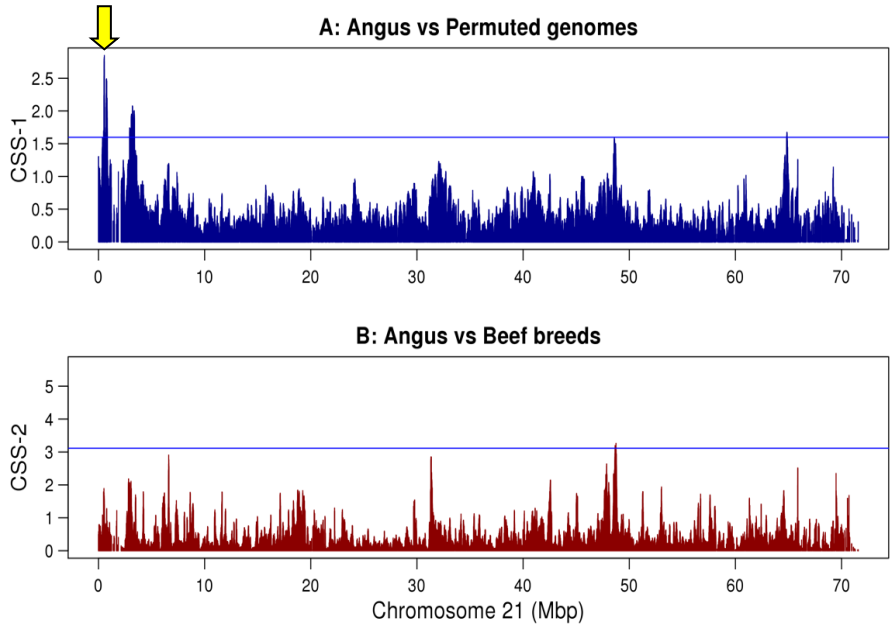

Figure 3. Composite selection signals on chromosome 21 in Angus cattle, computed by using two different reference populations: A) Permuted genomes, B) Beef breeds. Blue lines at genome-wide top 0.1%. Arrow at the top shows position of the expected signatures of selection in Angus

On BTA14 (24-27 Mbp of UMD3.1 bovine genome assembly), a strong selective sweep has been established due to selection targeting beneficial variants in *PLAG1-CHCHD7* in Angus and many other breeds (Randhawa *et al.* 2016). On BTA16 (42-48 Mbp), an extensive region under selection has been found underlying several genes important for both dairy and beef production. At the start of BTA21 (0.3-4 Mbp), many breeds (Angus, Belgian Blue, Brahman, Holstein) have shown strong selection signatures (Randhawa *et al.* 2016). Therefore, this study expected to find significant CSS on the target regions of BTA14, BTA16 and BTA21. CSS-1 succeeded in detecting these regions, while the genetic composition of reference population composed of beef breeds in CCS-2 showed a major confounding influence on detection of selection signatures in Angus (Figures 1-3). Interestingly, several genomic regions, e.g., BTA1, BTA4 and BTA13, also strongly selected in Angus were captured by both CSS-1 and CSS-2 (results not shown). This suggested that only the genomic patterns of allele frequencies and haplotype structures which are relatively more similar in a candidate breed and reference population can neutralise the across-breed statistics of selection signatures. Moreover, the magnitude of CSS values also varied between CSS-1 (max: 2.85, top 0.1%: 1.6) and CSS-2 (max: 5.67, top 0.1%: 3.1). The circular permutation approach was initially proposed to decide on a significance threshold for the empirical genome-wide association (Cabrera *et al.* 2012) and selection signatures (Stainton *et al.* 2015). The genome-wide results suggest that using the maximum value of CSS-1 as a significance threshold for conventional CSS-2 approach may have recovered a few confounded regions. However, most of the expected regions did not show a cluster of high values in CSS-2 analyses. Thus, our new approach of using permuted genomes as the reference population has been proved advantageous. This approach can be used to detect selection signatures from single breed data by permuting from its own samples where data are limited, or a unique set of variants is available through whole-genome sequencing.

CONCLUSIONS

Molecular data can provide insights into historical natural or artificial selection events and the genetic architecture underlying breed-specific traits. This study examined the impact of reference population to validate known genomic regions under selection in Angus cattle. The results provide evidence that a reference population of closely related or phenotypically similar to the candidate breeds affects the power to detect selection signatures. Our new approach of circular genomic permutations can potentially limit such confounding effects and resource-limited data can be efficiently analysed to detect historic selection pressures in any species.

ACKNOWLEDGEMENTS

We thank Robert D. Schnabel and Jerry F. Taylor from University of Missouri, Columbia, USA for providing the genotypic data.

REFERENCES

Cabrera C.P., Navarro P., Huffman J.E., Wright A.F., Hayward C., Campbell H., Wilson J.F., Rudan I., Hastie N.D*., et al.* (2012) *G3-Genes Genom. Genet.* **2**:1067.

Gibbs R.A., Taylor J.F., Van Tassell C.P., Barendse W., Eversole K.A., Gill C.A., Green R.D., Hamernik D.L., Kappes S.M*., et al.* (2009) *Science* **324**:528.

Kemper K.E. and Goddard M.E. (2012) *Hum. Mol. Genet.* **21**:R45.

Randhawa I.A.S., Khatkar M.S., Thomson P.C. and Raadsma H.W. (2014) *BMC Genet.* **15**:34.

Randhawa I.A.S., Khatkar M.S., Thomson P.C. and Raadsma H.W. (2016) *PLoS ONE* **11**:e0153013.

Stainton J.J., Haley C.S., Charlesworth B., Kranis A., Watson K. and Wiener P. (2015) *Anim. Genet.* **46**:37.