**Application**

Characterisation of copy number variants has heretofore not been available on the Irish sheep population, this coupled with knowledge of their impact on animal performance may lead to greater genetic improvement for the Irish sheep industry.

**Introduction**

The advent of low cost DNA genotyping for ruminants with the primary objective of improving genetic evaluations and providing accurate parentage has resulted in the genotyping of large numbers of animals (including sheep). However, a multitude of additional uses for high density genotypes exist; one such application is copy number variant (CNV) detection. A CNV is a structural change in DNA, normally defined to be > 1kb in length, where a region of DNA is duplicated or deleted. Copy number variants have been extensively studied in cattle with associations observed across a multitude of production traits. In comparison there is a paucity of such studies in sheep; with a small number of studies showing an association between CNVs and milk and growth traits (Di Gerlando et al., 2019; Ladeira et al., 2022). However, these studies focused on a single breed and used a relatively low density single-nucleotide polymorphism (SNP) panel, with a limited number of studies on sheep using a higher density SNP panel for CNV detection. Due to cost restraints, the lower density panels tend to be more widely used for sheep and no previous study has compared the role of panel density on CNV detection. The objective of the current study was to characterise and investigate the functional impact of CNVs in a multi-breed sheep population comprising of eight breeds. A secondary objective was to validate the impact of panel density on CNV identification.

**Materials and Methods**

Genotype data, including genotyping intensity values, was available for 174 Irish sheep from 8 breeds genotyped on the Illumina OvineSNP600K BeadChip platform. Only genotypes from animals with a SNP call rate of ≥0.9, and SNPs with a call rate ≥0.9 were retained for analysis. Only autosomal SNPs with a known chromosome and position were retained. After edits, 561,891 SNPs on 174 animals were available for analysis. To investigate the capacity of low density SNP panels to identify CNVs, a SNP panel resembling the Ovine SNP50 panel with 41,458 autosomal SNPs common to both the OvineSNP600K panel and the OvineSNP50K was generated for each animal, this panel will be hereafter referred to as the SNP50 panel. For both the SNP50K and SNP600K panels, CNVs were called using an integrated hidden Markov model within the PennCNV software (Wang et al., 2007). Only CNVs with a length of 3 or more SNPs were retained across both panels. To remove potential false positive CNV calls, animals with a log R ratio (LRR) standard deviation > 0.3 , B allele frequency (BAF) drift > 0.01, or absolute waviness factor > 0.05 were removed from further analysis (37 animals were removed). A CNV was determined to be common to an animal across both genotyping panels if the CNV was detected on both the SNP50K and SNP600K panel. Overlapping CNVs that overlapped by at least one base pair were grouped into regions known as CNVRs using the software HandyCNV and were classified into three categories: deletions, duplications or mixed (containing at least one deletion and one duplication). CNVRs were generated both across all breeds and within breed. Sheep QTLdb was then used to identify CNVRs that are associated with economically important traits.

**Results**

A total 5,201 CNVs were called from the SNP600K panel across 137 animals, consisting of 3,447 deletions and 1,754 duplications. Across animals, there was a median of 13 deletions and 9 duplications detected per animal. Deletions had a mean length of 66.15 kb and duplications had a mean length of 55.11 kb. In comparison only 244 CNVs were called from the SNP50 panel comprising 146 deletions and 98 duplications. The distribution of CNVs was found to differ across breeds, which may be due to their variations in origin and breed management. The mean length of CNVs called from the SNP50 panel was greater than those called by the SNP600K panel for both deletions and duplications. There was a median of 2 deletions and 1 duplication per animal using the SNP50 panel. Only 25% of CNVs called by the SNP50 panel overlapped with a CNV called by the SNP600K panel and had a mean length of 267.31 kb whereas CNVs with no overlap had a mean length of 140.17kb. Copy number variants called from the SNP600K panel were grouped into 2041 CNVR. The most common CNVR, found in 66 animals, was located on chromosome 10, between 71,051,498 and 72,205,564 bp and has been shown to overlap with quantitative trait loci (QTL) associated with carcass traits. In total there were 131 CNVR found in more than 5% of animals.

**Conclusions**

Copy number variants are a common feature of the sheep genome that are not accounted for in breeding programmes. The distribution of CNVs varies among different breeds, and these CNV regions have been found to overlap with areas containing genes associated with various functional traits. Low density panels, such as the SNP50K, have limited capacity to accurately identify CNVs; they typically identify longer CNVs and tend to have a high false discovery rate.

**References**

Di Gerlando, R., Sutera, A. M., Mastrangelo, S., Tolone, M., Portolano, B., Sottile, G., Bagnato, A., Strillacci, M. G., & Sardina, M. T. (2019). Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. PLOS ONE, 14(4), e0215204.

Ladeira, G. C., Pilonetto, F., Fernandes, A. C., Bóscollo, P. P., Dauria, B. D., Titto, C. G., Coutinho, L. L., e Silva, F. F., Pinto, L. F. B., & Mourão, G. B. (2022). CNV detection and their association with growth, efficiency and carcass traits in Santa Inês sheep. Journal of Animal Breeding and Genetics, 139(4), 476–487.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. A., Hakonarson, H., & Bucan, M. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Research, 17(11), 1665–1674.