

Integrating copy number variants to improve the resolution of genome-wide association analyses for milk production traits in Croatian Istrian sheep

C. Rochus<sup>1</sup>, M. Špehar<sup>2</sup>, A. Kasap<sup>3</sup>, Z. Barač<sup>4</sup>, D. Ceranac<sup>2</sup>, J. Ramljak<sup>3</sup>, I. Pocrnić<sup>1</sup>

<sup>1</sup> University of Edinburgh, Roslin Institute, Easter Bush Campus, EH25 9RG Midlothian, United Kingdom, <sup>2</sup> Agency for Agriculture and Food, Svetošimunska 25, 10000 Zagreb, Croatia, <sup>3</sup> University of Zagreb, Faculty of Agriculture, Svetošimunska 25, 10000 Zagreb, Croatia, <sup>4</sup> Ministry of Agriculture, Grada Vukovara 78, 10000 Zagreb, Croatia

Genome-wide association studies (GWAS) help to characterise the genetic architecture of traits. GWAS often use the most common type of genomic variants, single nucleotide polymorphisms (SNP) because they are easy and affordable to identify and genotype. Copy number variants (CNV) are structural variants that are deletions or insertions in the genome larger than 50 base pairs in size. CNV are particularly interesting in the GWAS context, as they can affect gene structure, expression, and regulation, potentially capturing genetic variation missed by SNP. While more challenging to detect and genotype, CNV can be identified from SNP arrays (using total signal intensity and allelic intensity ratio data) or whole-genome sequence (WGS) data. We propose using SNP and CNV in GWAS to dissect the genetics of milk production traits (milk, fat and protein yields, and somatic cell score) in Istrian sheep. The transnational Istrian sheep breed is a good example of a locally adapted sheep breed, crucial for the livelihood of extensive small-scale dairy farmers, and a reservoir of genetic diversity with unique characteristics needed to meet future challenges for livestock production systems. We have used pedigree data for registered Croatian Istrian sheep, 41,170 test day milk records, 50K SNP array data for 1,265 animals, and short-read WGS data (~30x coverage) for 15 key animals. We have identified CNV in genotyped and sequenced animals, imputed SNP and CNV to include in the GWAS and compared the results with standard GWAS methods. We have done functional gene annotation in regions of interest by leveraging available information from published public datasets. Local breeds, like Istrian sheep, are important for maintaining species-level genetic diversity, to be able to respond to future challenges, and understanding the genetic architecture of their traits through GWAS will contribute to this.