Detection of interchromosomal rearrangements in bulls using large genotype and phenotype datasets

J Jourdain, H Barasc, T Faraut, C Grohs, C Donnadieu, A Pinton, D Boichard, A Capitan

Detection of interchromosomal rearrangements in bulls using large genotype and phenotype datasets

J. Jourdain[1,2], H. Barasc[3], T. Faraut[3], C. Grohs[1], C. Donnadieu[4], A. Pinton[3], D. Boichard[1] and A. Capitan[1,2]
[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, G2B, Domaine de Vilvert, 78350 Jouy en Josas, France, [2]Eliance, 149, Rue de Bercy, 75012 PARIS, France, [3]GenPhySE, Université de Toulouse, INRAE, ENVT, 23 Chemin des Capelles, 31320 Castanet-Tolosan, France, [4]INRAE, US 1426, Université Fédérale de Toulouse, GeT-PlaGe, Genotoul, France Génomique, 24 chemin de borde rouge, Auzeville, 31326 Castanet-Tolosan, France; jeanlin.jourdain@inrae.fr

Interchromosomal rearrangements (IR), which result from the transfer of genetic material from one chromosome to another, can have severe phenotypic consequences due to gene dosage defects. Artificial insemination (AI) bulls are not currently screened for IR before their semen is used. Most IR have been detected by surveillance for the t(1;29) Robertsonian fusion or by targeted controls of low fertility bulls. Here, we developed a method to detect IR using linkage disequilibrium (LD) across chromosomes and applied it to 5,571 paternal halfsib families genotyped for genomic evaluation. Thirteen progeny groups (0.23%) showed significant LD, and 12 were confirmed by cytogenetic analyses: one Robertsonian fusion, 10 reciprocal translocations, and the first case of insertional translocation reported in cattle. Using national databases, reciprocal translocation carriers were all found in the worst percentile of their breed for male fertility and their carrier daughters were subfertile. The insertional translocation carrier was the worst bull for mortality with 44% of daughters dying before 365 days of age. We used long-read sequences for 7 bulls to characterize breakpoints and detect genes putatively affected in their expression, possibly leading to deviating phenotypes in balanced progeny, such as the observed delayed growth and high death rate in some balanced daughters. In addition to systematic karyotyping of bulls to avoid IR diffusion, our study highlights the importance of computing LD in halfsib groups with our method to detect remaining smaller IR and to manage carriers in the population. This study is the most comprehensive scan of the cattle population scan and paves the way for follow-up studies on the origins and consequences of IR. JJ is a recipient of a CIFRE PhD grant with the financial support of ANRT and APIS-GENE.

Functional information embedded in the unmapped short reads of whole-genome sequencing

G.B. Neumann[1], P. Korkuć[1], M. Reißmann[1], M.J. Wolf[2], K. May[2], S. König[2] and G.A. Brockmann[1]
[1]Humboldt-Universität zu Berlin, Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences, Invalidenstrasse 42, Ostbau, 10115 Berlin, Germany, [2]Justus-Liebig-Universität, Institute of Animal Breeding and Genetics, Ludwigstr. 21, 35390 Gießen, Germany; guilherme.neumann@hu-berlin.de

Livestock genomics involves resequencing genetic information from additional individuals of the same species with an established reference genome. Short-read sequencing generates DNA sequences of 150-300 bp, which are mapped to the reference genome. Unmapped reads are typically discarded, but they may contain information about pathogenic DNA and structural variants (SVs) not present in the reference genome. In this study, we analysed unmapped reads from whole-genome sequencing of 302 German Black Pied cattle (DSN) to explore these hypotheses. The unmapped reads retrieved from the 302 DSN animals were assembled into scaffolds and blasted against the NCBI's database for reference and representative genomes of all available species. Scaffolds mapping against those genomes covering at least 10% of their respective reference genomes were kept for further analysis. SVs were detected for both mapped and unmapped short reads using Delly and SvABA, respectively. Among the unmapped reads of 302 sequenced DSN animals, 116 contained assembled DNA sequences covering >10% of the genome of eleven different species of bacteria and six viruses. Of those species, Mycoplasma and bovine parvovirus 3 are known pathogens infecting cattle. DNA sequences covering more than 10% of foreign eukaryotes were not found in the assembled unmapped reads. All 302 DSN animals had assembled unmapped reads aligned to Bos species with an average length of 341.3 kb per animal. Using unmapped reads, 26,593 SVs were detected in addition to 19,410 SVs that had been detected with mapped reads. While metagenomics is typically used to detect pathogens, unmapped reads of whole-genome sequencing can also provide valuable information on the presence of pathogens which is crucial for epidemiologists. Furthermore, the examination of unmapped reads revealed additional SVs that could not be identified through the alignment to the *Bos taurus* reference genome. This emphasizes the need for high-quality long sequence reads to accurately detect SVs.