The construction of a cattle pangenome for 14 French dairy and beef breeds provides new insights into their genetic diversity.

Valentin Sorin, Marie-Pierre Sanchez, Carole Iampietro, Camille Eché, Amandine Suin, Camille Marcuzzo, Laurence Drouilhet, Denis Milan, Gwenola Tosser-Klopp, Cécile Donnadieu, et al.

Session 15 Theatre 2

Construction of a cattle pangenome for 14 French dairy and beef breeds provides new insights into their genetic diversity

V. Sorin[1], M. P. Sanchez[1], C. Iampietro[2], C. Eché[2], A. Suin[2], C. Marcuzzo[2], L. Drouilhet[3], D. Milan[2,3], G. Tosser-Klopp[3], C. Donnadieu[2], C. Gaspin[4], C. Birbes[5], C. Klopp[5], D. Boichard[1], M. Boussaha[1]

[1] Université Paris Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France, [2] INRAE, US 1426, GeT-PlaGe, Genotoul, France Genomique, Universite Fédérale de Toulouse, 31326 Castanet-Tolosan, France, [3] Université de Toulouse, INRAE, GenPhySE, Castanet-Tolosan, 31326 Castanet-Tolosan, France, [4] Genotoul Bioinfo, BioInfoMics, MIAT UR875, INRAE, Castanet-Tolosan, 31326 Castanet-Tolosan, France, [5] Sigenae, Genotoul Bioinfo, BioInfoMics, MIAT UR875, INRAE, Castanet-Tolosan, 31326 Castanet-Tolosan, France

The current cattle genome reference assembly, based on a single Hereford cow fails to capture the whole spectrum of genetic variations within the species. Structural variations (SVs), defined as genomic variations longer than 50 nucleotides, can have a potential impact on both complex and Mendelian phenotypic variations. However they are difficult to detect using only standard approaches of either short or long-read sequence mapping to the current bovine genome assembly. Thanks to the recent advances in long-read sequencing technologies coupled with the development of appropriate bioinformatics tools, it's now possible to construct de novo genome assemblies for a large number of animals across various cattle breeds. It also offers the opportunity to study a broader range of both small and more complex genome-wide variations. Using these technologies, we have produced a comprehensive cattle pangenome incorporating genetic diversity from 64 high-quality de novo assemblies representing 14 French bovine dairy and beef breeds. We applied a combination of complementary approaches to characterize a wide spectrum of SVs and we report the identification of several megabases of novel genome sequences that are absent in the current cattle genome reference assembly. Further work is currently in progress to investigate the gene content of these non-reference sequences. This work was conducted in the SeqOccIn project, funded by the Occitanie region, FEDER, and Apis-Gene. Valentin Sorin's PhD is supported by INRAE.

Session 15 Theatre 3

Construction and characterization of a comprehensive ovine pangenome from 11 breeds provides new insights into their genetic diversity

V. Sorin[1], C. Birbes[2], C. Eché[3], C. Marcuzzo[3], J. Sarry[4], A. Suin[3], C. Donnadieu[3], C. Gaspin[5], C. Iampietro[3], D. Milan[3,4], C. Klopp[2], M. Boussaha[1], L. Drouilhet[4], M. P. Sanchez[1], G. Tosser-Klopp[4]

[1] Université Paris Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France, [2] Sigenae, Genotoul Bioinfo, BioInfoMics, MIAT UR875, INRAE, 24 Chem. de Borde Rouge, 31326 Castanet-Tolosan, France, [3] INRAE, US 1426, GeT-PlaGe, Genotoul, France Genomique, Universite de Toulouse, 24 Chem. de Borde Rouge, 31326 Castanet-Tolosan, France, [4] Université de Toulouse, INRAE, GenPhySE, 24 Chem. de Borde Rouge, 31326 Castanet-Tolosan, France, [5] Genotoul Bioinfo, BioInfoMics, MIAT UR875, INRAE, 24 Chem. de Borde Rouge, 31326 Castanet-Tolosan, France

The current method of constructing a genome reference assembly through sequencing the entire genome of just one or a very small number of individuals falls short in capturing the complete spectrum of genetic diversity within most species, including ovine. In addition, the exploration of genomic structural variations (> 50 nucleotides) is constrained when relying solely on short-read sequences aligned to these genome reference sequences. Nevertheless, due to the rapid advancements in sequencing technologies and bioinformatics tools, it is now possible to generate long-read sequences and create de novo genome assemblies for numerous animals, thereby enabling the commencement of pangenomic studies within these species. In this study, we integrate long and short read sequences to build de novo assemblies and haplotype-resolved assemblies for 11 distinct ovine breeds. Employing various methodologies, we established a pangenome and subsequently investigated a wide range of structural variations within the species. We identified several additional megabases of genome sequences absent in the current reference genome assembly and we are currently investigating gene content of these non-reference sequences. This research was conducted as part of the SeqOccIn project, funded by the Occitanie region, FEDER, and Apis-Gene. Valentin Sorin's PhD is supported by INRAE.