HAL
open science

The 1000 bull genomes project - Toward genomic selection from whole genome sequence data in dairy and beef cattle

Benjamin Hayes, Hans D. Daetwyler, Ruedi Fries, Bernt Guldbrandtsen, Mogens Sando Lund, Didier Boichard, Paul Stothard, Roel F. Veerkamp, Ina Hulsegge, Dominique Rocha, et al.

To cite this version:

Benjamin Hayes, Hans D. Daetwyler, Ruedi Fries, Bernt Guldbrandtsen, Mogens Sando Lund, et al.. The 1000 bull genomes project - Toward genomic selection from whole genome sequence data in dairy and beef cattle. International Plant & Animal Genome XXI, Jan 2013, San Diego, CA, United States. hal-01019053

HAL Id: hal-01019053

https://hal.science/hal-01019053v1

Submitted on 6 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Benjamin Hayes *Biosciences Research Division, DPI*, **Hans D. Daetwyler** *Department of Primary Industries (Victoria)*, **Ruedi Fries** *Chair of Animal Breeding - Technische Universität München*, **Bernt Guldbrandtsen** *Aarhus University*, **Mogens Sando Lund** *Aarhus University, Dept of Genetics and Biotechnology*, **Didier A. Boichard** *INRA UMR1313 Animal Genetics and Integrative Biology*, **Paul Stothard** *University of Alberta, Edmonton*, **Roel F. Veerkamp** *Wageningen UR Livestock Research*, **Ina Hulsegge** *Animal Breeding and Genomics Centre*, **Dominique Rocha** *INRA UMR1313 Animal Genetics and Integrative Biology*, **Curtis Van Tassell** *USDA-ARS-BFGL*, **Erik Mullaart** *CRV*, **Birgit Gredler** *QualitasAG, Switzerland*, **Tom Druet** *Unit of Animal Genomics, GIGA , University of Liege*, **Alessandro Bagnato** *Università degli Studi di Milano*, **Mike Goddard** *Biosciences Research Division, DPI*, **Amanda Chamberlain** *Department of Primary Industries, Bundoora, Victoria*, **1000 Bull Genomes Consortium** *Multiple Institutions*

W150 - The 1000 Bull Genomes Project - Toward Genomic Selection From Whole Genome Sequence Data In Dairy and Beef Cattle

Genomic prediction of breeding values is now used as the basis for selection of dairy cattle, and in some cases beef cattle, in a number of countries. When genomic prediction was introduced most of the information was to thought to be derived from linkage disequilibrium between markers and causative variants. It has become clear that much of the predictive power, based on 50,000 DNA markers, in fact derives from prediction of the effect of large chromosome segments that segregate within fairly closely related animals. This has lead to problems with across breed prediction, rapid decay of predictive power over generations and insufficient accuracy in some situations. Using full genome sequence data in genomic prediction should overcome these problems. If linkage disequilibrium between SNP on standard arrays and causative mutations affecting the quantitative trait is incomplete, accuracy of prediction should be improved as a result of including the actual causative mutations affecting the trait of interest in the data set. Secondly, persistence of accuracy of genomic predictions across generations will be improved with full sequence data, as the genomic predictions no longer depend on associations between SNP and causative mutations which currently erode quite rapidly with recombination. Thirdly, if genomic predictions are made across breeds, using full sequence data is likely to be particularly advantageous, as there is no longer a need to rely on marker- associations which may not persist across breeds. However, the cost of sequencing is such that the very large numbers of animals required for genomic prediction will not be sequenced An alternative strategy is to sequence key ancestors of the population, then impute the genotypes for the sequence variants into much larger reference sets with phenotypes and SNP panel genotypes. The 1000 Bull Genomes Project aims at building such a resource of sequenced key ancestor bulls for the bovine research community. The most recent run of the project included 238 full genome sequences of 130 Holstein, 43 Fleckvieh, 48 Angus and 15 Jersey bulls, sequenced at an average of 10.5 fold coverage. There were 25.2 million filtered sequence variants detected in the sequences, including 23.5 million SNP and 1.7 million insertion-deletions. Agreement of sequence genotypes to genotypes from an 800K SNP array in the sequenced Holstein bulls, where there was most data, was excellent at 98.8%. This increased to 99.7% when the genotypes were imputed based on all sequences. Concordance was slightly lower in other breeds. This project will provide an excellent opportunity to identify the most important causative variants, leading to greater understanding of biology underlying quantitative traits. Examples are given of genomic predictions for fertility, health and production traits using imputed sequence data.