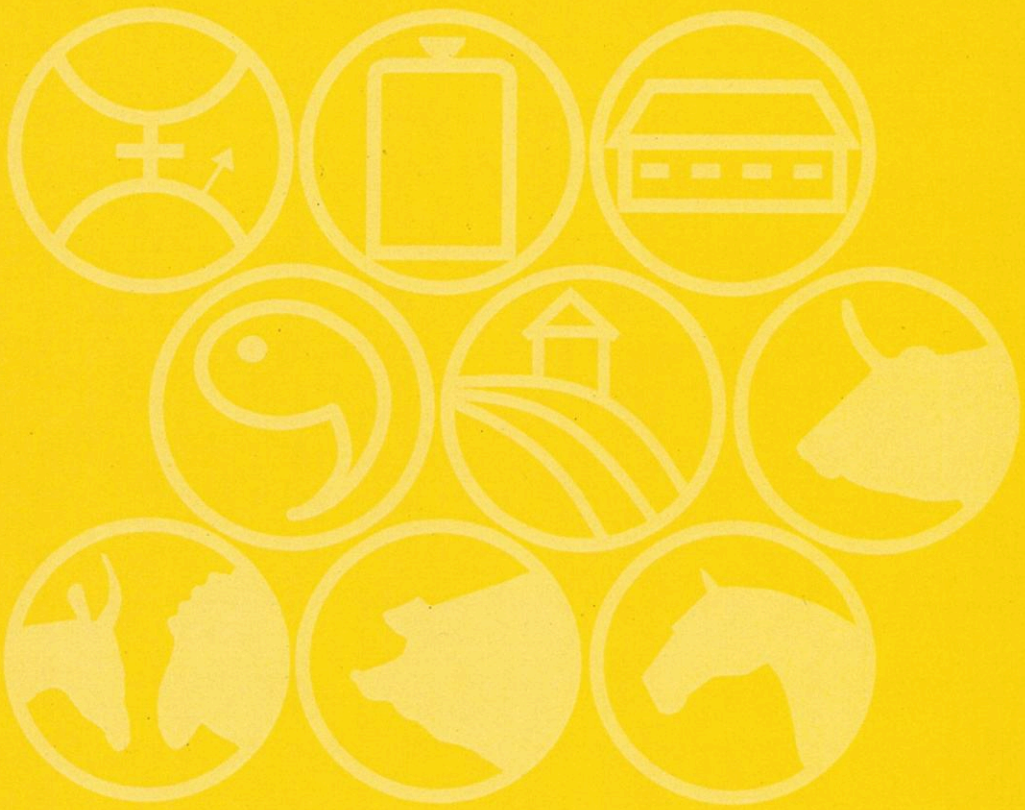
Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science



Book of abstracts No. 18 (2012)
Bratislava, Slovakia
27 - 31 August 2012

Error rate for imputation from BovineSNP50 to BovineHD

Schrooten, C.[1], Dassonneville, R.[2], Brøndum, R.[3], Chen, J.[4], Liu, Z.[4] and Druet, T.[5], [1]CRV BV, P.O. Box 454, 6800 AL Arnhem, Netherlands, [2]INRA, UMR GABI, Domaine de Vilvert, 78352 Jouy-en-Josas, France, [3]Aarhus University, Department of Molecular Biology and Genetics, Blichers Allé 20, 8830 Tjele, Denmark, [4]VIT, Heideweg 1, 27283 Verden, Germany, [5]Unit of Animal Genomics, Faculty of Veterinary Medicine / CBIG, University of Liège, 4000 Liège, Belgium; chris.schrooten@crv4all.com

Starting in 2007, large numbers of cattle have been genotyped with 50k platforms, mainly the BovineSNP50 BeadChip, to obtain more reliable breeding values for selection of animals. Since then, higher density genotyping platforms have been developed, among others the Illumina BovineHD BeadChip with 777k SNP. All animals could be regenotyped with the BovineHD chip, but the expected gain in reliability is expected not to outweigh the additional cost. Therefore, genotyping part of the population with the BovineHD and subsequently imputing the animals genotyped with 50k chips to HD is considered an attractive alternative, provided that imputation can be done at low error rate. The objective of this study was to investigate the error rate for imputation from BovineSNP50 genotypes to BovineHD. BovineHD genotypes were obtained for 548 high impact bulls from the Eurogenomics reference population. Genotypes for all but the BovineSNP50 markers were masked in 60 validation animals, different for each of 4 subsets. These 60 animals did not have descendants with BovineHD genotypes. The BovineHD genotypes of the remaining 488 animals in each subset were used as reference to impute the masked genotypes of the 60 animals. Imputation errors were computed as the number of differences between imputed alleles and observed alleles, divided by the number of compared alleles. When Beagle 3.3 was used, imputation error ranged from 0.55% to 0.76%. Combining Beagle with DAGPHASE did not improve results, indicating no benefit from adding linkage information with DAGPHASE. It was concluded that imputation from 50k platforms to HD can be done at low error rate when using Beagle 3.3.

SNPpit: efficient data management for high density genotyping

Groeneveld, E. and Truong, C.V.C., Institute of Farm Animal Genetics, Mariensee (FLI), Höltystr 10, 31535 Neustadt, Germany; eildert.groeneveld@fli.bund.de

Expanding use of high density genotyping creates substantial data management problems, both in terms of storage requirements as well as computational demands arising from data manipulation. Given the background of ever increasing panel densities across all species and reducing genotyping costs, SNPpit has been developed with the objective to store, manage, and retrieve genotype data of different SNP panels in a uniform way in one database, while leaving the data analysis to other packages like PLINK and GenABEL. Thus, SNPpit can be used as a central repository of all genotyping data accumulating within a lab. Highly compressed internal storage of genotype data allows to efficiently store large volumes of data. As a result, a 500 GB hard disk will be able to hold 100 billion genotypes which is equivalent to 200 panels of 500K genotyped for a total of 200,000 individuals. SNPpit provides for the creation of named subsets of genotypes by viewing them as matrices, with one of their dimensions being defined through a selection list of SNPs and the other a list of individuals. Through their names, SNPpit can export each subset for further analysis by software packages like PLINK. Because of their unique way of definition, derived datasets – as they may arise during editing or by focussing on certain chromosomes – can be stored practically without space requirements as only the SNP and individual selection vectors need to be saved in the database. Any combination of an existing SNP and individual selection vector will then result in a new genotype set, which can be exported for further use. In this way, disk space issues of derived datasets are eliminated. Because of the efficient compressed storage scheme data import and export are fast: importing 20 mio genotypes on 80 individuals took 39 sec while the export finished in 14 sec on an Intel Core i5 laptop. SNPpit is written in Perl using the PostgreSQL database for mass data storage. As all software components are released under the GPL, installations can be made freely.