Predicting homozygosity-by-descent to manage inbreeding and diversity

N.S. Forneris[1], T. Druet[1]

[1]Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Avenue de l'Hôpital, 1, 4000 Liège, Belgium; nforneris@uliege.be

Prediction of homozygous-by-descent (HBD) segments in the future offspring of a given mating is a valuable tool for the management of inbreeding and diversity as it provides information on expected inbreeding levels, at both genome-wide and local scales, and has proven to be efficient. Such prediction amounts to identify identical-by-descent (IBD) segments in the parents. Here we took advantage of 98 whole-genome sequenced trios (sire-dam-offspring) of a Holstein pedigree (264 animals, 84 couples) to evaluate different HBD prediction methods that we classified as follows : a) hidden Markov models (HMM) that model IBD along the four parental chromosomes, with 15, 9 or 3 IBD classes; b) ruled-based haplotype-matching methods that scan for identity-by-state segments; c) methods designed to estimate inbreeding within individuals (ZooRoH or the rule-based ROH) applied on pairs of parental haplotypes; d) SNP-by-SNP measures; and e) pedigree-based. Medium-density (50K) genotypes from the parents were used to predict inbreeding in the offspring. Scenarios with lower marker density data, genotyping-by-sequencing (GBS) data or smaller sample sizes were also investigated. The methods were compared in terms of correlations between predicted and reference genome-wide inbreeding levels or between locus-specific values (HBD probabilities). Reference inbreeding levels were defined with respect to different based populations (recent versus ancient inbreeding). Two HMM approaches (relying respectively on a 15 IBD class model and on ZooRoH) proved accurate, flexible and robust to predict recent inbreeding across the different tested scenarios, both at the genome-wide ($r = 0.74$-0.75) and locus-specific levels ($r = 0.48$-0.49). An efficient haplotype-matching method proved fast and accurate for genome-wide prediction. Use of GBS (both globally and locally) and lower marker density (at the locus-specific level) data resulted in lower performances than those obtained with the medium density array. Lower sample size did not affect the relative behavior of the predictors, when results were averaged across subsets. We are currently running simulations, with two demographic scenarios, to further characterize the different methods.