GenTORE

Genomic management Tools to Optimise Resilience and Efficiency

Grant agreement n°: 727213

H2020 - Research and Innovation Action

D4.1
Report about genomic predictions for purebred and crossbred individuals

---

Due date: M42 (November 2020)

Actual submission date: M42 (November 2020)

Project start date: 1st June 2017   **Duration:** 60 months

Workpackage concerned: WP4

Concerned workpackage leader: Mogens Sandø Lund

Lead Beneficiary: AU

Dissemination level:

   X PU: Public (must be available on the website)

   ☐ **CO:** Confidential, only for members of the consortium (including the Commission Services)

   ☐ **CI:** Classified, as referred to in Commission Decision 2001/844/EC

Table of content

1. Summary

A genomic prediction model, which can handle data from admixed individuals together with data from pure breeds, was developed and tested with simulated data. The model relies on breed origin information of each allele. Scenarios tested to validate our model included different levels of trait heritability, genetic correlation between populations and various forms of reference population structure. We also tested if our model can efficiently use local genomic correlations between the breeds to further improve prediction accuracies. Our results showed that the use of newly developed model improves genomic predictions due to increased population size for each breed as it facilitates the inclusion of data from admixed individuals. This increase in population size is more beneficial to small breeds. Our model also led to improved accuracies for admixed individuals, in some cases up to ten percentage points, which is a large improvement.

2. Introduction

Breeding programs for dairy cattle have been successful in improving production traits within breeds, but at the cost of deterioration of functional traits, and increased rate of inbreeding (Sørensen et al., 2008). Crossbreeding emerges as an efficient strategy for dairy cattle breeding to overcome these issues, and to achieve better productivity and performance. In such schemes, where crossbreeding is used, for instance in ProCROSS system (http://www.procross.info), crossbred dams are mated to purebred sires from different pure-breeds, resulting in genetic composition of crossbred animals becoming admixture of the pure breeds included in the rotation. As a result, admixture proportions of individuals vary at each rotation cycle due to the choice of sire breed (Crum et al., 2019). It is, therefore, required in some breeding programs that genomic prediction models are able to accommodate a reference population including admixed individuals, as well as multiple pure breeds, allowing simultaneous evaluation of all selection candidates. Traditionally, data of admixed individuals are not used, they neither get any genomic evaluation.

If a large number of commercial farm data for admixed populations becomes available, it can help to improve selection accuracy by: (i) expanding the data size for each pure breed population, (ii) providing phenotypes which are not available for purebred animals, and (iii) allowing the inclusion of heterotic effects in selection decisions (Veroneze et al., 2014). How to use those data in genomic evaluations is still an open question. Naturally, all purebred and admixed individual data can be combined together, when SNP effects are assumed to be the same across the breeds.

An appealing approach to make use of data of admixed individuals in genomic prediction is to incorporate breed proportions in genomic prediction models. Those breed proportions are traditionally computed based on pedigree or genome-wide markers. However, approaches relying on a single value of breed proportions for each individual are suboptimum, as two individuals with exactly the same breed proportions may have very different patterns of admixture over their genome depending on which chromosomal region is inherited from which pure breed. Methods relying on breed origin of each SNP (chromosomal segment) can explicitly account for such pattern of admixture over an individual's genome.

In task 4.1, we investigated methodologies for genomic prediction using a reference population of multiple purebred and admixed individuals. A genomic prediction method relying breed origin of each allele was developed, and tested against the traditional method that include breed proportions. Using simulations, we investigated the accuracy of genomic prediction using different approaches: (i) treating the data as of a single homogeneous

population, (ii) considering breed-specific SNP effects w/o accounting for correlations between the breeds, and (iii) using priors that lead region specific correlations among the breeds.

3. Results

Adding admixed population's (MIX) data to the combined pure breed data as of a different breed led to higher accuracies than using combined pure breed data alone. When prediction models were able to account for breed origin of alleles, accuracies were generally higher than those from combining all available data. This was dependent on the correlation of QTL effects between the breeds. Accuracies varied when using SNP effects from any of the pure breeds to predict the breeding values of MIX, reflecting their most recent ancestor breed. Using the breed-specific SNP effects estimated separately in each pure breed, while accounting for breed origin of alleles for the selection candidates of MIX, accuracies were generally improved compared to using SNP effects from multiple pure breed's reference data. This was dependent on the correlation of QTL effects between the breeds. Models able to accommodate MIX data with breed origin of alleles approach generally led to higher accuracies than those from the models without considering breed origin of alleles, and assuming a single joint population of all animals. Analysis assuming correlated SNP effects between the breeds had no advantage over the analysis assuming uncorrelated SNP effects.

4. Conclusions

Our results show that the use of admixed individuals' data together with pure breeds' data in genomic prediction while considering breed origin of alleles has two main advantages. First, it increases the data size for all pure breeds, allowing more accurate estimation of breeding values for all breeds, particularly for the breed with a small population size. Second, it increases the prediction accuracy for admixed individuals. This will enable farmers to make more informed breeding decisions for his cows.

5. Partners involved in the work

AU, ALLICE

6. Annexes

References

Crum, T., Schnabel, R., Decker, J., Regitano, L., Taylor, J.: CRUMBLER: A tool for the prediction of ancestry in cattle. PLoS ONE 14(8), 0221471 (2019)

Sørensen, M., Norberg, E., Pedersen, J., Christensen, L.: Invited review: Crossbreeding in dairy cattle: A Danish perspective. J Dairy Sci 91(11), 4116–4128 (2008)

Veroneze, R., Bastiaansen, J., Knol, E., Guimaraes, S., Silva, F., Harlizius, B., Lopes, M.S., Lopes, P.: Linkage disequilibrium patterns and persistence of phase in purebred and crossbred pig (Sus scrofa) populations. BMC Genet 15(1), 126 (2014)