200. The effect of recombination rate on genomic selection in simulation

M. Johnsson[1*], A.M. Johansson[1], T. Klingström[1] and E. Rius-Vilarrasa[2*]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; [2]Växa Sverige, P.O. Box 30204, 104 25 Stockholm, Sweden; martin.johnsson@slu.se

Abstract

Genome features such as recombination rate vary between different regions of the genome and affect the distribution of genetic variation, thus potentially affecting the response of the genome to selection and the accuracy of genomic selection. In this paper, we build a quantitative genetic simulation with variable recombination rate based on real linkage maps. It confirms that higher recombination rate causes a somewhat higher long-term genetic gain due to preservation of additive genetic variance, and that high recombination rate lowers the accuracy of genomic selection. While the benefit happens in the long term, the decreased accuracy is immediate.

Introduction

In this paper, we build a quantitative genetic simulation with variable recombination rate in order to explore the effect of recombination rate on genomic selection in simulation. Features of the genome such as recombination rate, mutation rate and historical natural selection vary along the genome, and shape the distribution of genetic variation as they affect the supply of standing variation (Corbett-Detig *et al.* 2015, Harpak *et al.* 2016) and the extent of selective interference between variants (Hill and Robertson 1966).

Simulations are widely used in animal genetics to test the performance of analysis methods, understand the response to selection in breeding program designs, but also for mechanistic modelling of genetic variation. These simulations make many simplifying assumptions, including, often, abstracting away genome features and treating the genome as a more or less uniform grid of markers. Some of these assumptions may be entirely warranted, relying on the generality of quantitative genetic theory, but others may not.

In our research on the genome dynamics of livestock breeding, we hypothesise that there is additional knowledge about genetic architecture, and potentially genomic selection accuracy, to be gained by modelling the relationship between variation and features of the genome in more detail. This paper presents some of our pilot work on the effect of recombination rate on selection, using published physical and genetic maps of the cattle and chicken genomes.

Materials & methods

We used AlphaSimR (Gaynor, Gorjanc and Hickey 2021) to build a simulation of a closed breeding population with different karyotypes and recombination rate variation based on real linkage maps from cattle and chickens.

Simulation of population and traits. The founder population was generated using the coalescent simulator MaCS within AlphaSimR, using the 'GENERIC' population history, which models a large historical population size decreasing during domestication, and a final effective population size of 1000. We simulated one quantitative trait with either 100, 1000 or 10,000 causative variants drawn at random, proportioned to chromosomes based on their physical length. Additive genetic coefficients were drawn from a standard normal distribution. Dominance degrees were drawn from a normal distribution with mean 0.2 and variance 0.1, similar to values estimated by Bennewitz and Meuwissen (2010). Normally distributed noise was added to the phenotypes to achieve a heritability of 0.4.

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_200, © M. Johnsson *et al.* 2022

Chromosomes and linkage maps. Chromosomes were either based on the cattle karyotype, the chicken karyotype or were 20 chromosomes of equal size. Table 1 describes the different genomes simulated, with number of chromosomes, physical and genetic sizes. We only included autosomes that were included in the linkage map, meaning that the X and Z chromosome are missing, respectively, as well as some chicken microchromosomes. We used linkage maps from Ma *et al.* (2015) for cattle and Elferink *et al.* (2010) for the chicken. In order to update the physical coordinates, we used the UCSC LiftOver service to sequentially lift the physical position to ARS-UCD1.2 for cattle and GRCg6a for the chicken. We filtered the lifted maps to remove markers that caused disagreements in marker order, leaving a total of 58,010 in cattle and 12,907 in the chicken. We adjusted the genetic maps from coalescent simulation, which have uniform recombination rate, to correspond to the real linkage maps before breeding. We converted the initial genetic position of each site to a physical position, and linearly interpolated the new genetic position on the real linkage map based on that physical position. Linkage maps and code are available at https://github.com/mrtnj/lifted_recombination_maps.

Breeding structure. Each generation, 50 males (10%) and 100 females (20%) were selected to be the parents of next generation, and mated to produce 1000 offspring, half of each sex. The first ten generations of each replicate used phenotypic selection, and for the next ten generations, we either ran another ten generations of phenotypic selection or ten generations of genomic selection. In order to measure the decline in genomic accuracy, training was performed once using data from generation 6-9. The genomic selection model was SNP-BLUP based on 50,000 SNP markers randomly chosen but equally proportioned to chromosomes based on physical length. We evaluated accuracy as the correlation between true and the estimated breeding value. Each simulation case was run for 50 replicates. The simulation code is available at https://github.com/mrtnj/wcgalp_recombination.

Simulation with linkage equilibrium. For comparison, we created a simulation where the causative variants had the same starting frequencies, and distributions of additive and dominance coefficients, but where the loci were made to be in linkage equilibrium independently sampling the genotypes from a binomial distribution each generation. The linkage equilibrium simulation also used the same heritability, population size and a sex-averaged selection intensity as the AlphaSimR simulation.

Results

High recombination rate led to a lower loss of additive genetic variance on average (Figure 1). This effect was noticeable when there were many causative variants (i.e. not in the case with 100 loci), at later generations, but the difference was small compared to the variation between replicates. This was accompanied by somewhat higher genetic mean. Considering the case with 10,000 loci, in generation 20, the genetic mean was 4.4% higher for 1 M chromosomes, 6.4% higher for 2 M chromosomes, 3.6% higher for the cattle genome and 4.0% higher for the chicken genome compared to the shortest genome (0.5 M chromosomes).

Table 1. Genomes simulated.

Case	Physical length (Mbp)	Genetic length (M)	Chromosomes
0.5 M chromosomes	2,000	10	20
1 M chromosomes	2,000	20	20
2 M chromosomes	2,000	40	20
Cattle genome	2,489	24	29
Chicken genome	941	27	27

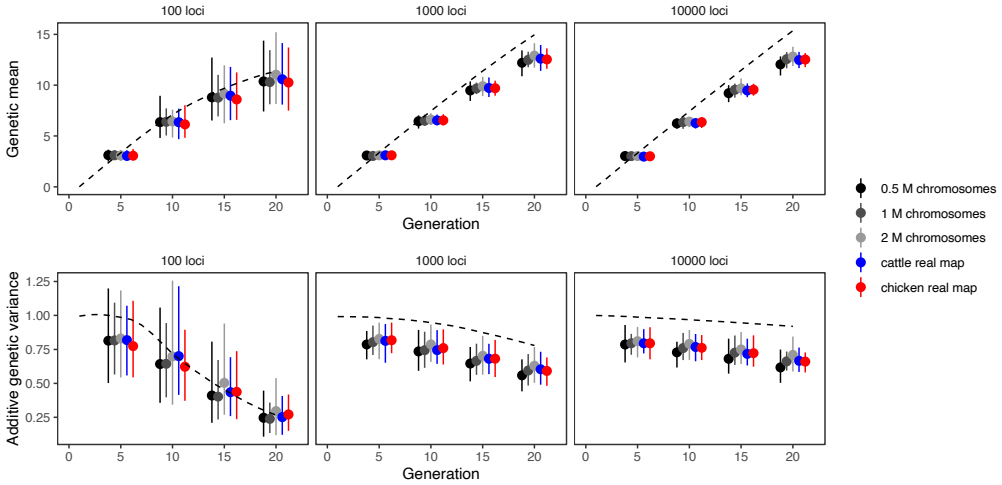

Figure 1. Changes in additive genetic variance with different recombination rates. The dashed line is the average from a simulation where causative variants are in linkage equilibrium.

On the other hand, higher recombination rate caused a decrease in genomic prediction accuracy (Figure 2). Considering the case with 10,000 loci, the first generation of genomic selection, average accuracies after one generation were 19% lower for 1 M chromosomes, 36% lower for 2 M chromosomes, 25% lower for the cattle genome, and 22% for the chicken genome compared to the most accurate 0.5 M chromosome case. When the genomes used the real cattle or chicken karyotype but had uniform recombination rate, the reduction was somewhat greater than with real linkage maps (27% for cattle and 26% for the chicken).

Discussion

The slight increase in genetic gain due to higher recombination is consistent with results from Battagin *et al.* (2011), who found that higher recombination rate led to higher genetic gain due to a smaller reduction in additive genetic variance, though the effect was small. The quantitative results are not comparable, as they used a different breeding structure, and selection on true breeding value. With genomic selection, the benefit from high recombination rate would be offset by a decreased genomic selection accuracy; while the improvement due to higher recombination happens in the long term and is small, the decrease in accuracy is immediate.

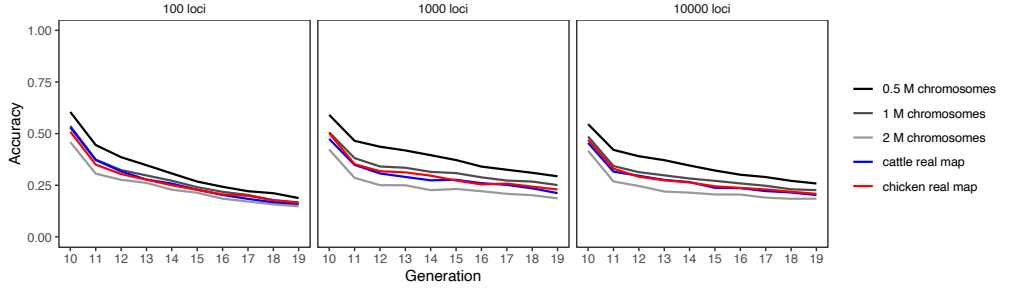

Figure 2. Decline in genomic selection accuracy over generations increasingly removed from a constant reference population, with different recombination rates.

As there is substantial variation within the genome in recombination rate, different regions of the genome should display local variation in its contribution to genomic selection accuracy, and to a lesser extent to long term gain. In the linkage maps used here, the 25 and 75% quantiles of recombination rate in between-marker intervals were 0.060 and 1.6 cM/Mbp for the cattle linkage map and 0.0 and 4.1 cM/Mbp for chicken linkage map.

Except for the real linkage maps, most aspects of these simulations were highly simplified. These are some avenues for development: the simulations used a generic population history, with a domestication bottleneck and large ancestral population. Work is ongoing to infer population history, and synthesise that with published estimates. As the real linkage maps were only introduced after the coalescent simulation of the founder population, the regional variation in linkage disequilibrium was only built up during the simulated breeding. Also, the real recombination rate landscape is more punctuated, with short 'hotspots' of a few kilobasepairs. Work is ongoing to infer fine-scale recombination rate and use it to improve the realism of the founder simulation. Finally, we used a generic breeding structure of a small closed population. The flexibility of AlphaSimR as well as recent developments to communicate and visualise breeding structures (Simianer *et al.* 2021) will be helpful for improving this aspect.

Acknowledgements

The research was supported by grant Dnr. 2020-01637 from Formas – a Swedish Research Council for sustainable development.

References

Battagin, M., Gorjanc, G., Faux, A. M., Johnston, S. E., & Hickey, J. M. (2016) Genet Sel Evol 48(1) https://doi.org/10.1186/s12711-016-0221-1

Bennewitz, J., and Meuwissen, T. H. E. (2010) J Anim Breed Genet 127(3), 171-179 https://doi.org/10.1111/j.1439-0388.2009.00847.x

Corbett-Detig, R. B., Hartl, D. L., and Sackton, T. B. (2015) PLoS Biol 13(4):e1002112. https://doi.org/10.1371/journal.pbio.1002112

Elferink, M. G., van As, P., Veenendaal, T., Crooijmans, R. P., and Groenen, M. A. (2010) BMC Genet 11(1) https://doi.org/10.1186/1471-2156-11-11

Gaynor, R. C., Gorjanc, G., and Hickey, J. M. (2021) G3 11(2):jkaa017 https://doi.org/10.1093/g3journal/jkaa017

Harpak, A., Bhaskar, A., and Pritchard, J. K. (2016) PLoS Genet, 12(12):e1006489. https://doi.org/10.1371/journal.pgen.1006489

Hill, W. G., and Robertson, A. (1966) Genetics Res 8(3):269-294 https://doi.org/10.1017/S0016672300010156

Ma, L., O'Connell, J. R., VanRaden, P. M., Shen, B., Padhi, A., *et al.* (2015). PLoS Genet 11(11): e1005387 https://doi.org/10.1371/journal.pgen.1005387

Simianer, H., Büttgen, L., Ganesan, A., Ha, N. T., and Pook, T. (2021) J Anim Breed Genet 138(2), 137-150 https://doi.org/10.1111/jbg.12534