# 199. The long-term effects of genomic selection: allele frequency changes and fixation of loci

Y.C.J. Wientjes[1*], P. Bijma[1], J. van den Heuvel[1], B.J. Zwaan[1], Z.G. Vitezica[2] and M.P.L. Calus[1]

*[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]INRAE, UMR 1388 GenPhySE, 31326 Castanet-Tolosan, France; yvonne.wientjes@wur.nl*

## Abstract

Allele frequencies change due to drift and selection. We investigated those changes for different selection methods. We simulated a livestock population with 50 generations of selection for an additive trait controlled by 2,000 segregating loci at the start of selection. Our results show that the average absolute change in allele frequency with genomic selection was slightly higher than with pedigree selection, and considerably higher than with mass selection. Genomic and pedigree selection both resulted in fixation of roughly three times more loci than mass selection, and roughly five times more loci became fixed for the unfavourable allele. While this was mainly a result of increased genetic drift, genomic selection also lost additional favourable alleles due to hitchhiking. This suggests that genomic selection can limit long-term genetic gain.

## Introduction

Despite the sometimes strong selection pressure in animal breeding programs, genetic variation and rates of genetic gain have been stable for many generations in several animal populations (Beniwal *et al.*, 1992; Havenstein *et al.*, 2003). Therefore, these pedigree-based selection methods have proven to be sustainable so far.

Implementation of genomic selection has increased the rates of genetic gain in the short term. Moreover, it has likely accelerated the changes in allele frequencies across generations in certain regions on the genome (Heidaritabar *et al.*, 2014; Doekes *et al.*, 2018) and has increased the risk of losing rare favourable alleles (Jannink, 2010; Liu *et al.*, 2014). However, the long-term consequences of genomic selection are still largely unknown.

The current availability of genomic information has also opened up the possibility to investigate the changes in allele frequency as a result of selection, which is essential for understanding the long-term effects of selection. The vast majority of causal loci are, however, unknown. Therefore, our aim is to investigate and quantify the allele frequency changes at causal loci and how this is affected by selection methods using simulations.

## Materials & methods

We simulated a livestock population under 50 generations of selection. First, we simulated a historical population in QMSim software (Sargolzaei and Schenkel, 2009) that showed a linkage disequilibrium pattern and allele frequency distribution comparable to livestock populations. Then we used our own developed Fortran program to generate the next 50 generations of selection. In every generation, the best 100 females and 100 males were selected and randomly mated using a mating ratio of 1:1 and a litter size of 10 (5 females and 5 males). Each scenario was replicated 20 times. For more details, see Wientjes *et al.* (2021).

**Genome and phenotypes.** The simulated genome had 10 chromosomes of 100 cM each. At the end of the historical population, 20,000 segregating loci with a uniform allele frequency distribution were selected as marker, and 2,000 segregating loci with a U-shaped allele frequency distribution were selected as causal

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


DOI: 10.3920/978-90-8686-940-4_199, © Y.C.J. Wientjes *et al.* 2022

loci with an effect on the simulated trait. Moreover, 4,000 non-segregating loci were randomly selected as locations for causal mutations, which started to affect the trait when the locus was segregating after a mutation took place. The number of new mutations for each individual was sampled from a Poisson distribution, with an average of 0.6, which resulted in a mutational variance of ~$0.001\sigma_e^2$ per generation.

Phenotypes were simulated based on an additive model. For all causal loci and mutations, an additive effect was sampled from $N(0,1)$. The additive effects and genotypic information of causal loci were used to calculate the true breeding value for each individual. Finally, a non-genetic residual was added to the true breeding value, to obtain a heritability of 0.4.

**Selection process.** We compared five methods to select the parents of the next generation:

1. RANDOM, which randomly selected the animals.
2. MASS, which selected the animals with the best phenotypes.
3. PBLUP_OP, which selected the animals with the best pedigree based estimated breeding value, estimated including own performance information.
4. GBLUP_NoOP, which selected the animals with the best genomic estimated breeding value (GEBV), estimated excluding own performance information.
5. GBLUP_OP, which selected the animals with the best GEBV, estimated including own performance information.

With PBLUP, the last eight generations of pedigree and last three generations of phenotypes were used. With GBLUP, the last three generations of genotypes and phenotypes were used. Breeding values were estimated with the MTG2 software (Lee and van der Werf, 2016), using a model including a fixed mean, a random additive genetic effect and a residual.

**Allele frequency changes and fixation of loci.** We compared the average change in allele frequency and the number of loci that became fixed during the 50 generations of selection, requiring a minimal change in allele frequency of 0.2 to disregard loci that initially were already close to fixation. For the loci that became fixed, we investigated the average additive effect and starting allele frequency and whether the favourable alleles was fixed or lost.

To disentangle the impact of drift and selection, additional simulations were run with random selection using a population with the same effective population size ($N_e$) as in the selection scenarios. For each scenario, the $N_e$ was estimated using the pedigree kinship coefficient based on the off-diagonal elements of the pedigree relationship matrix. Then we measured the effect of selection as the additional number of loci that became fixed in the selection scenarios compared to the drift scenario.

## Results

All scenarios with selection resulted in an improvement of the phenotype over the 50 generations. This improvement was largest for GBLUP_OP and lowest for GBLUP_NoOP (Table 1). Both GBLUP methods lost ~82% of the genetic variance, which considerably reduced the rate of genetic gain across generations. The loss in genetic variance was considerably less with MASS (~60%). Therefore, the rate of genetic gain after 50 generations was highest for MASS. The average change in allele frequency was 2-3 times larger with selection than with RANDOM, and largest for the GBLUP scenarios, followed by PBLUP_OP and finally MASS. Compared to RANDOM ($N_e$=222), selection reduced $N_e$ most for PBLUP_OP (46), followed by the GBLUP scenarios (74-82), and finally MASS (136).

**Table 1.** Change in average phenotype, genetic variance and allele frequency across 50 generations of selection[1].

| | Phenotypic change[2] | | Proportion of genetic variance lost | | Average change in allele frequency | | Effective population size | |
|---|---|---|---|---|---|---|---|---|
| RANDOM | 0.02 | (0.11) | 0.03 | (0.03) | 0.023 | (0.000) | 222 | (0.8) |
| MASS | 27.31 | (0.44) | 0.60 | (0.01) | 0.054 | (0.001) | 136 | (1.8) |
| PBLUP_OP | 27.02 | (0.46) | 0.79 | (0.01) | 0.062 | (0.000) | 46 | (0.9) |
| GBLUP_NoOP | 25.25 | (0.45) | 0.82 | (0.01) | 0.063 | (0.001) | 74 | (1.4) |
| GBLUP_OP | 29.37 | (0.42) | 0.83 | (0.00) | 0.065 | (0.001) | 82 | (1.5) |

[1] Results are shown as averages across replicates with standard errors across replicates between brackets.

[2] Expressed in base generation additive genetic standard deviations.

With RANDOM, very few loci became fixed, and there was an equal chance of fixing or losing the favourable allele (Table 2). Selection resulted in fixation of a considerable number of loci, being similar for PBLUP_OP, GBLUP_NoOP and GBLUP_OP, which was ~2.7 times higher than with MASS. With selection, more loci became fixed for the favourable allele than for the unfavourable allele, which was more emphasized with MASS (~89%) than with the other selection scenarios (~80%).

With selection, the average additive effect of the loci where the favourable allele became fixed was larger than for an average locus (average α across all loci was 0.041), and was largest with MASS. The average starting allele frequency of those loci was lower with selection, and lowest for GBLUP_OP, GBLUP_NoOP and PBLUP_OP. In contrast, the loci where the unfavourable allele became fixed had on average a lower additive effect than an average locus (0.024 vs 0.041). Moreover, the average starting allele frequency at those loci was largest for GBLUP_OP, GBLUP_NoOP, PBLUP_OP, followed by MASS and lowest for RANDOM.

For all selection methods, more loci became fixed for the favourable allele due to selection than due to drift, while more loci became fixed for the unfavourable allele due to drift. Both GBLUP methods and PBLUP_OP fixed a similar number of loci, but the impact of selection on the fixation of loci was higher for the GBLUP methods. Moreover, with both GBLUP methods, an additional number of loci became fixed for the unfavourable allele as a result of selection.

**Table 2.** Characteristics of loci that became fixed over 50 generations of selection[1].

| | | | Favourable allele | | | Unfavourable allele | | |
|---|---|---|---|---|---|---|---|---|
| | Total nr.[2] | Prop. fav.[3] | Avg. α[4] | Start freq.[5] | Sel.[6] | Avg. α[4] | Start freq.[5] | Sel.[6] |
| Random | 2 (0.4) | 0.46 (0.09) | 0.035 | 0.74 | 0.00 | 0.046 | 0.25 | 0.00 |
| MASS | 133 (5.3) | 0.89 (0.01) | 0.061 | 0.59 | 0.18 | 0.021 | 0.34 | 0.02 |
| PBLUP_OP | 358 (4.6) | 0.80 (0.01) | 0.053 | 0.52 | 0.25 | 0.025 | 0.39 | 0.01 |
| GBLUP_NoOP | 356 (5.7) | 0.79 (0.00) | 0.052 | 0.52 | 0.30 | 0.027 | 0.40 | 0.07 |
| GBLUP_OP | 357 (5.3) | 0.82 (0.00) | 0.054 | 0.51 | 0.31 | 0.024 | 0.40 | 0.06 |

[1] Results are shown as averages across replicates with standard errors of the mean between brackets.

[2] Considering loci with a minimum change in allele frequency of 0.2.

[3] Proportion of loci fixed for the favourable allele.

[4] Average absolute statistical additive effect (α) in generation 0 in genetic standard deviation units.

[5] Frequency of the favourable allele in generation 0.

[6] Proportion of loci segregating in generation 0 that became fixed due to selection.

## Discussion

The loss in genetic variance was considerably larger with the BLUP methods than with MASS, and MASS had the highest rate of genetic gain in the last generations. Therefore, for the longer term, MASS is expected to outperform GBLUP_OP in terms of cumulative genetic gain. The BLUP methods also showed a faster change in allele frequency and a greater rate of fixation of favourable alleles than MASS, which can be explained by the higher accuracy of the BLUP methods (Liu *et al.*, 2014).

The $N_e$ was lowest for PBLUP_OP, which is a result of more family selection with PBLUP compared to GBLUP (Liu *et al.*, 2014; Wientjes *et al.*, 2021). So even though the average change in allele frequency and total number of loci becoming fixed was similar for GBLUP and PBLUP_OP, this was more due to drift and less due to selection with PBLUP_OP (21% due to drift and 25% selection) than with GBLUP (15% due to drift and 31% due to selection).

With the two GBLUP scenarios, an additional number of unfavourable alleles became fixed over and above the effect of drift. This was surprising, since selection was expected to prevent loss of favourable alleles on top of the effect of drift. An additional loss of favourable alleles due to selection did not happen with MASS and PBLUP, indicating that GBLUP resulted in greater fixation of unfavourable alleles due to genetic hitchhiking (Smith and Haigh, 1974; Barton, 2000).

Overall, GBLUP resulted in slightly larger and faster changes in allele frequencies of causal loci than PBLUP, and much larger and faster changes than MASS. As a consequence, GBLUP lost more favourable alleles. Maintaining rare favourable alleles is especially important for long-term genetic gain (Jannink, 2010), suggesting that GBLUP can limit the long-term genetic gain and should be accompanied by a management of genetic diversity.

## Acknowledgements

This research is supported by the Netherlands Organisation of Scientific Research (NWO; VENI grant 16774). The use of the HPC cluster has been made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR).

## References

Barton N.H. (2000) Phil. Trans. R. Soc. B 355(1403): 1553-1562. https://doi.org/10.1098/rstb.2000.0716

Beniwal B.K., Hastings I.M., Thompson R., and Hill W.G. (1992) Heredity 69(4): 361-371. https://doi.org/10.1038/hdy.1992.136

Doekes H.P., Veerkamp R.F., Bijma P., Hiemstra S.J., and Windig J.J. (2018) Genet. Sel. Evol. 50(1): 15. https://doi.org/10.1186/s12711-018-0385-y

Havenstein G.B., Ferket P.R., and Qureshi M.A. (2003) Poult. Sci. 82(10): 1500-1508. https://doi.org/10.1093/ps/82.10.1500

Heidaritabar M., Vereijken A., Muir W.M., Meuwissen T., Cheng H. *et al.* (2014) Heredity 113(6): 503-513. https://doi.org/10.1038/hdy.2014.55

Jannink J.-L. (2010) Genet. Sel. Evol. 42(1): 35. https://doi.org/10.1186/1297-9686-42-35

Lee S.H., and van der Werf J.H.J. (2016) Bioinformatics 32(9): 1420-1422. https://doi.org/10.1093/bioinformatics/btw012

Liu H., Sørensen A.C., Meuwissen T.H.E., and Berg P. (2014) Genet. Sel. Evol. 46(1): 8. https://doi.org/10.1186/1297-9686-46-8

Sargolzaei M., and Schenkel F.S. (2009) Bioinformatics 25(5): 680-681. https://doi.org/10.1093 /bioinformatics/btp045

Smith J.M., and Haigh J. (1974) Genet. Res. 23(1): 23-35. https://doi.org/10.1017/S0016672300014634

Wientjes Y.C.J., Bijma P., Calus M.P.L., Zwaan B.J., Vitezica Z.G. *et al.* (2021) BioRxiv: 2021.2003.2016.435664. https://doi.org/10.1101/2021.03.16.435664