Long-term changes in genetic mean and genic variance and underlying allele frequencies in breeding programmes

G.A. de Oliveira Junior[1], M. Spehar[2], F.S. Schenkel[1], C.F. Baes[1], and G. Gorjanc[2]

[1] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada; [2] The Roslin Institute, University of Edinburgh, Edinburgh, UK; [*]gregor.gorjanc@roslin.ed.ac.uk

Abstract

The infinitesimal model assumes that polygenic traits are influenced by (infinitely) many quantitative trait loci with (infinitely) small effects and that selection on such traits has minimal impact on the frequency of alleles at individual loci, hence minimal impact on genetic variance. Here we evaluated trends in genetic mean and genic variance for milk yield over 25 years in 9,403 proven and genotyped bulls from the Canadian Dairy Network. The results showed significant change in genetic mean and minor change in genic variance, supporting the infinitesimal model. A detailed analysis of SNP marker allele frequencies indicated sizeable changes at individual loci. We grouped the SNP markers by their initial allele frequency and direction of change, to indicate how these groups of makers contribute to changes in mean and variance in the analysed period of 25 years and predictions for the next 75 years showing sizeable changes in mean and variance.

Introduction

Animal breeding programmes focus on additive genetic values that represent the sum of allele substitution effects over all quantitative trait loci for the trait of interest. Much of the theory and estimation of additive genetic values is based on the infinitesimal model. The infinitesimal model assumes that polygenic traits are influenced by (infinitely) many quantitative trait loci with (infinitely) small effects (Fisher, 1918). Consequently, this model implies that selection on polygenic traits has (infinitely) small impact on allele frequency changes at individual loci (Fisher, 1918). Despite the (infinitely) small impact on allele frequency changes, selection can have a significant cumulative change across all loci, as manifested by breeding programmes.

Since the infinitesimal model predicts (infinitely) small allele frequency changes under selection, this implies (infinitely) small short-term changes in genetic variance. This prediction is in line with observations that response to selection in most breeding programmes are not levelling off (e.g., Hill, 2016). There are reports that recent introduction of genomic selection has sped up the rate of change in genetic mean and variance (Jannink, 2010; Hidalgo et al., 2020; Wientjes et al., 2021), which could also be due to the Bulmer effect, negative linkage-disequilibrium component of genetic variance (Bulmer, 1971; Lara et al., 2021).

Abundant genome-wide marker data is enabling rapid genetic improvement via genomic selection, but also detailed dissection of underlying genome changes that drive the rapid genetic improvement. Compared to phenotypic selection, genomic selection is increasing the risk of losing favourable alleles (Jannink, 2010; Wientjes et al., 2021), particularly for the rare alleles or those alleles that are in unfavourable linkage-disequilibrium with genomic selection markers. A systematic evaluation of the change in past and predicted future allele frequencies and their relation to genetic mean and variance is lacking and the aim of this contribution.

Materials & Methods

Data. We used data from the routine milk yield genetic evaluation of the Canadian Dairy Network – Lactanet (Guelph, ON). The data comprised 9,403 proven bulls, born between 1989 and 2014. We retrieved de-regressed estimated breeding values and associated reliabilities for the bulls. We also retrieved their single nucleotide polymorphism (SNP) marker genotypes. We retained 40,448 SNP markers from the autosomes having minor allele frequency above 0.05.

Estimation of marker effects. We estimated allele substitution effects of the SNP markers using the Bayesian ridge-regression model $\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{M\alpha} + \boldsymbol{e}$, where $\boldsymbol{y}$ is a vector of de-regressed estimated breeding values for milk yield scaled to mean zero and unit variance, $\boldsymbol{b}$ is a vector of fixed effects (the intercept), $\boldsymbol{\alpha}$ is a vector of allele substitution effects, $\boldsymbol{e}$ is a vector of residuals, $\mathbf{X}$ is an incidence matrix linking $\boldsymbol{y}$ with $\boldsymbol{b}$, and $\mathbf{M}$ is a matrix of SNP marker genotypes. The model assumptions were $\boldsymbol{\alpha} \sim N(\mathbf{0}, \boldsymbol{I}\sigma_\alpha^2)$ and $\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{W}\sigma_e^2)$, where $\mathbf{W}$ is a diagonal weight matrix with values equal to (1-Reliability$_i$)/Reliability$_i$ (for the *i*-th bull) adjusting heterogeneous residual variance in the de-regressed estimated breeding values.

We inferred the model parameters using the Gibbs sampler implemented in the JWAS package with default prior specifications (Cheng et al., 2018). We obtained 100,000 samples from posterior distribution of allele substitution effects and saved every 100-th sample. We also obtained posterior samples of breeding values ($\boldsymbol{a}$) for the *i*-th sample using $\boldsymbol{a^i} = \boldsymbol{M\alpha^i}$.

We further obtained samples from posterior distribution of the mean of breeding values $\mu_{a,k}^i = E(\boldsymbol{a}_k^i)$ and variance of breeding values $\sigma_{a,k}^{2,i} = Var(\boldsymbol{a}_k^i)$ for each (*k*-th) year of birth. Summarising the $\mu_{a,k}^i$ and $\sigma_{a,k}^{2,i}$ samples and plotting the summaries against the year gives time trends in genetic mean and variance. Lara et al. (2021) showed that changes in genetic variance are driven by changes in genic variance (function of allele frequencies and allele substitution effects at the causal loci; $\sigma_{a,l}^2 = 2p_l(1-p_l)\alpha_l^2$ for the *l*-th locus and $\sum_{l=1}^{n_l} \sigma_{a,l}^2$ for all n_l loci) and in linkage-disequilibrium covariance (function of correlation between causal locus genotypes and their allele substitution effects). Here we focus only on genic variance. Preliminary analysis showed minor changes in genic variance in the period between 1989 and 2014, despite significant changes in genetic mean, corroborating the infinitesimal model. To further analyse these results, we analysed changes in allele frequencies of the SNP markers.

Estimation of allele frequency time trends. We calculated allele frequency of the SNP markers for each year of birth (1989 to 2014) and estimated their rate of change with beta regression (*betareg* R package; Cribari-Neto and Zeileis, 2010), independently for each SNP marker. We have then predicted change in allele frequency from the fitted beta regression model for each SNP marker for 100 years ($\hat{p}_{l,k}$ for the *k*-th year), including the observed period (1989 to 2014).

To further analyse change in genetic mean and no apparent change in genic variance we have grouped the SNP markers according to the initial allele frequency (in year 1989) (above or below 0.5) and change in allele frequency (increasing or decreasing). The rational for this grouping is that following the $2p_l(1-p_l)$ expression we expect that "InitialAFAbove0.5IncreasingAF" markers will show decrease in genic variance, "InitialAFBelow0.5IncreasingAF" markers will show increase in genic variance until frequency of 0.5, "InitialAFAbove0.5DecreasingAF" markers will show increase in genic variance until frequency of 0.5, and "InitialAFBelow0.5DecreasingAF" markers will show decrease in genic variance. For each of the groups we plotted predicted allele frequencies and

showed their contribution to predicted genetic mean ($\hat{\mu}_{a,k} = \sum_{l=1}^{n_l} 2\hat{p}_{l,k}\hat{\alpha}_l$) and genic variance ($\left(\sum_{l=1}^{n_l} \hat{\sigma}_{a,l}^2\right)_k = \sum_{l=1}^{n_l} 2\hat{p}_{l,k}(1-\hat{p}_{l,k})\hat{\alpha}_l^2$), with $\hat{\alpha}_l$ being mean of the 1000 samples α_l^i.

Results

Figure 1 shows the predicted changes in allele frequency for the SNP markers over 100 years. The change of allele frequencies in each of the four groups of course match the grouping based on the estimated rate of change, but there is considerable variation within each group. Some SNP markers are predicted to fixate, but many allele frequency changes are small.

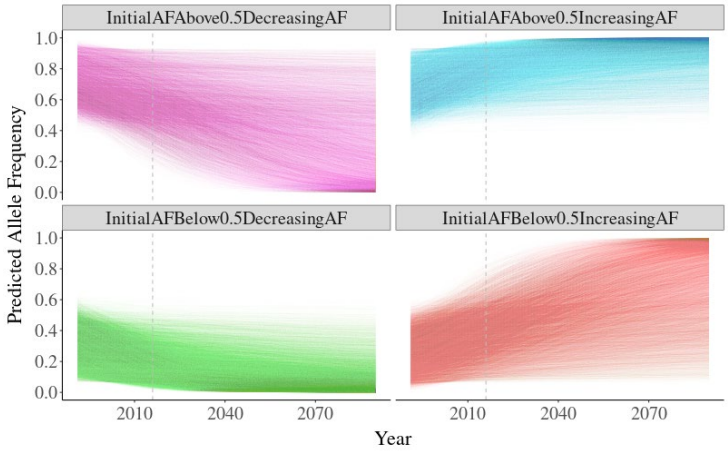

Figure 1. Predicted allele frequencies over 100 years for the four groups of SNP markers.

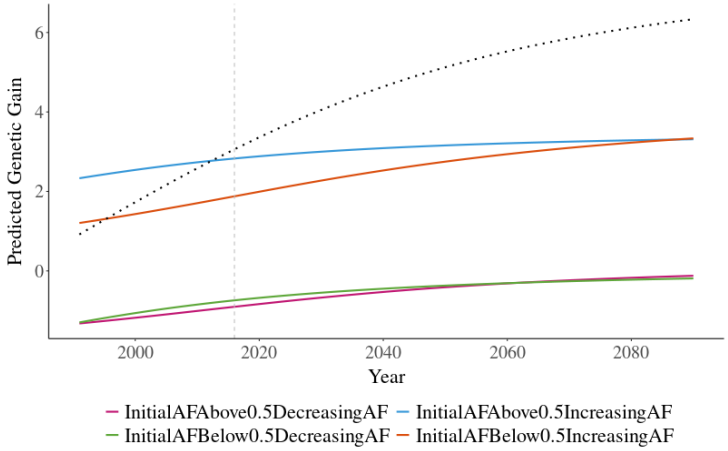

Figure 2. Predicted genetic mean over 100 years and contribution of the four groups of SNP markers to the genetic mean.

Figure 2 shows the predicted genetic mean over 100 years and contribution of the four groups of SNP markers to the genetic mean. Genetic mean increased over the initial period and is predicted to increase in the future, above 6 genetic standard deviations of genetic gain. The largest contribution is from the "InitialAFAbove0.5IncreasingAF" group, followed by the "InitialAFBelow0.5IncreasingAF" group, with a considerably smaller contribution of the other two groups of SNP markers.

Figure 3 shows the predicted genic variance over 100 years and contribution of the four groups of SNP markers to the genic variance. Genic variance was almost constant during the initial

period but is predicted to decrease for about a third in future. Contribution of the four groups varied over time, with relatively similar contributions in the initial period and the largest contribution of the "InitialAFBelow0.5IncreasingAF" and "InitialAFAbove0.5DecreasingAF" groups for most of the period, though these contributions decreased after some point. This result is in line with the expectation according to the $2p_l(1-p_l)$ expression.

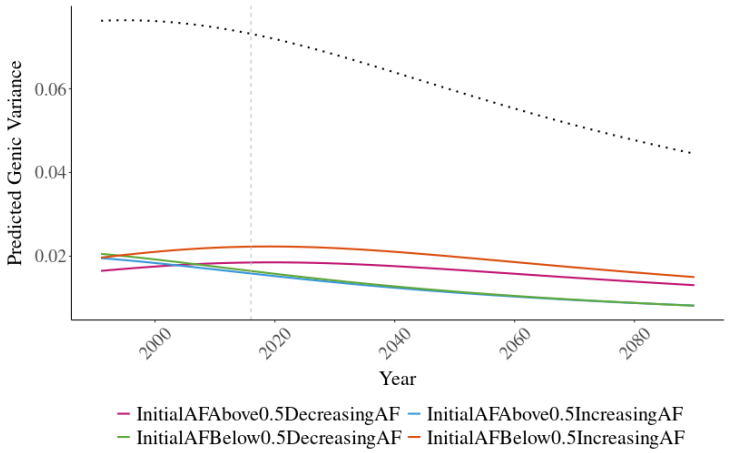

Figure 3. Predicted genic variance over 100 years and contribution of the four groups of SNP markers to the genic variance.

Discussion

We analysed change in genetic mean and variance over 25 years as well as underlying changes in SNP marker allele frequencies and predicted the trends in the same quantities for further 75 years. Results show that allele frequencies of the SNP markers changed sizeably during the analysed period of 25 years, and are predicted to change even more in the future, though arguably over a long period. The results suggests that there is considerable dynamics at individual loci that is not seen at the cumulative level. The results also give insights into the assumptions of the infinitesimal model and its predictions for changes in genic variance. The results also clearly showed different contribution of the four groups of SNP markers, both to genetic mean and genic variance. All these predictions assume that estimated rate of allele frequency changes in the initial period will hold in the future, that allele substitution effects will not change, hence that there are no non-additive genetic effects and that we have stable linkage-disequilibrium between the SNP markers and causal loci, and that there will be no mutations. These are clearly strong assumptions. The presented analysis nevertheless enables in-depth study of genetic changes in breeding populations.

References

Bulmer (1971) Am Nat 105:201–211. https://doi.org/10.1086/282718
Cheng et al. (2018) 11th WCGALP, Auckland, New Zealand.
Cribari-Neto and Zeileis (2010) J Stat Soft 34:1-24. https://doi.org/10.18637/jss.v034.i02
Fisher (1918) Proc Roy Soc 52:399–433. https://doi.org/10.1017/S0080456800012163
Hidalgo et. al. (2020) J Anim Sci 98:1–2. https://doi.org/10.1093/jas/skaa032
Hill (2016) Genetics 202:877–881. https://doi.org/10.1093/genetics/202.3.NP
Jannink (2010) Genet Sel Evol 42:35. https://doi.org/10.1186/1297-9686-42-35
Lara et al. (2021) Heredity 128:21–32. https://doi.org/10.1038/s41437-021-00485-y
Wientjes et al. (2021) bioRxiv 2021.03.16.435664. https://doi.org/10.1101/2021.03.16.435664