Inclusion of externally computed genotype frequencies in the gene-flow method

F. Ytournel* and H. Simianer
Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August University,
37075 Göttingen, Germany
* Corresponding author: fytourn@gwdg.de

INTRODUCTION

The evolution of the molecular and statistical methods provides the opportunity of a more precise knowledge of the positions of Quantitative Trait Loci. This knowledge is now routinely used in breeding nuclei for some traits (e.g. Leymarie et al. 2009). However, the outcomes in the overall population are not considered in this selection. The reference method to define a breeding program is the gene-flow method (Hill, 1974; Elsen and Mocquot, 1974). The central element of this method is the so-called P-matrix which describes the transmission paths between the different cohorts, classically defined by the purpose of the group, the sex and age of the animals. One thus follows in the whole population the genes from a given path over the generations. Börner and Reinsch (2009) recently proposed a way to account for imprinting in this matrix by subdividing each of the cohorts relatively to the paternal or maternal origin of the genes. We here more generally propose to consider each genotype as a subpart of the cohorts and account for the changes in the gene frequencies due to the selection in the breeding nucleus. These new frequencies were computed with a deterministic algorithm to optimize the genetic progress (Chakraborty *et al.*, 2002) adapted to overlapping generations (Ytournel and Simianer, 2009). The P-matrix thus changes from one generation to the next. We compare the results obtained (*a*) when keeping a constant P-matrix, (*b*) for various initial frequencies of the selected genes and (*c*) with selected and unselected breeding nuclei.

METHODS

Deterministic optimisation of the genetic progress with overlapping generations (Ytournel and Simianer 2009)

The method aims at maximizing the weighted sum of the genetic progress resulting from the selection on a genetic value including known genes over a definite time horizon (thereafter set to 15 generations). The individuals in the breeding nucleus were selected on a combination of an index value and two bi-allelic genes with effects 1.0 and 1.5 phenotypic standard deviations that were not linked. The male population (Breeding Males, BM) was composed of two age-classes (AC) while the female one (Breeding Females, BF) comprised three AC. The composition of the selection groups relatively to the age-classes is provided in Table 1. Two situations were considered: there is no selection or the male and female groups are subjected to selection with the proportions selected in each age-class as presented in Table 1. We considered the following initial frequencies of the most favourable haplotype: 0.01, 0.10, 0.25, 0.50, 0.75, 0.90 and 0.95. The genotype frequencies were recorded for each generation.

Table 1: Composition of the selection groups and selected proportions

	BM AC1	BM AC2	BF AC1	BF AC2	BF AC3
Proportion of the AC in the sex group	0.7	0.3	0.7	0.2	0.1
Selected proportion	0.3	0.8	0.5	0.8	0.5

Gene-flow

The population was composed of the two previous selection groups plus one of Production Females (PF) in which there were also 3 AC. The origin of the genes is described through the synthetic gene-flow matrix presented in Table 2.

From the previously obtained genotype frequencies, we computed the haplotype frequencies transmitted to the offspring generation which was made possible by distinguishing the parental origin of the haplotypes. For the PF, we supposed that the frequencies corresponded to the probabilities of obtaining each haplotype from a given genotype.

We followed the genes originating from the first genotype (the most favourable and thus the selected one) of the first male AC over the generations. We recorded the Standard Discounted Expressions (SDE) over the generations and the asymptotic value of the vector containing the gene proportions (m-vector) for the various scenarios in the breeding nucleus (selection and initial haplotype frequency). We also looked at the possible gain achieved by accounting for the variation of the genotype frequencies over the generations in the P-matrix.

Table 2: Synthetic P-matrix

		From		
		BM	BF	PF
To	BM	BM > BM	BF > BM	-
	BF	BM > BF	BF > BF	-
	PF	BM > PF	-	PF > PF

RESULTS

Influence of adapting the P-matrix to the changes in genotype frequencies

As expected, using a progressive P-matrix led to little difference when the breeding nucleus was not selected (Figures 1 and 2) but up to an increase in SDE of 0.052 in the selected population with an initial haplotype frequency of 0.5 (Figure 1). The proportional increase went up to 110% for an initial frequency of 0.1. The asymptotic values of the m-vector were completely confounded when no selection was applied, but were higher in selected populations when the P-matrix changed over the generations (Figure 2).

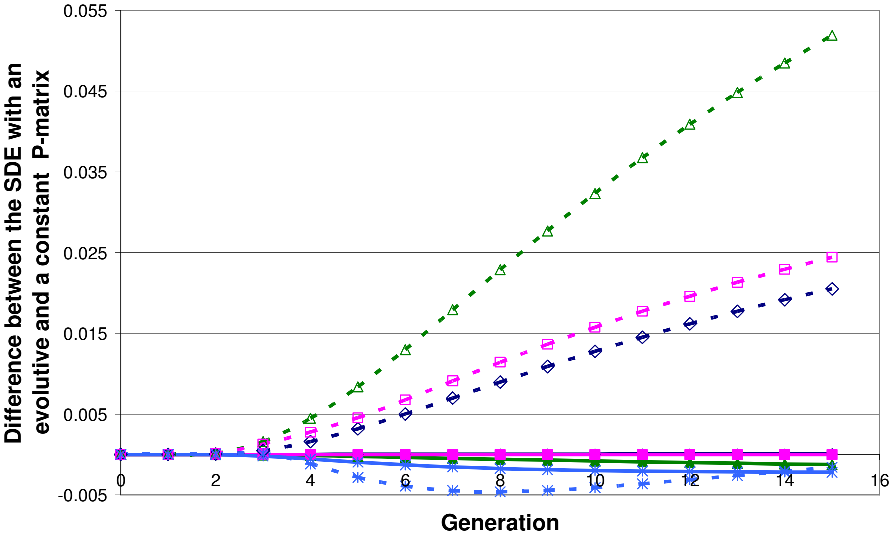

Figure 1: Difference in the SDE-values with a progressive or a constant P-matrix (Dotted lines: selected populations; solid lines: unselected populations. Initial frequencies: Diamonds: 0.1; Triangles: 0.5; Squares: 0.75; Stars: 0.9)

Influence of selection and initial haplotype frequencies

The proportion of the genes originating from the 1st genotype of the 1st male AC over the time horizon was higher when the breeding nucleus was subjected to selection and the initial frequency of the 1st haplotype was higher (Figure 2). The maximum difference was observed for an initial frequency of 0.5 with a progressive P-matrix.

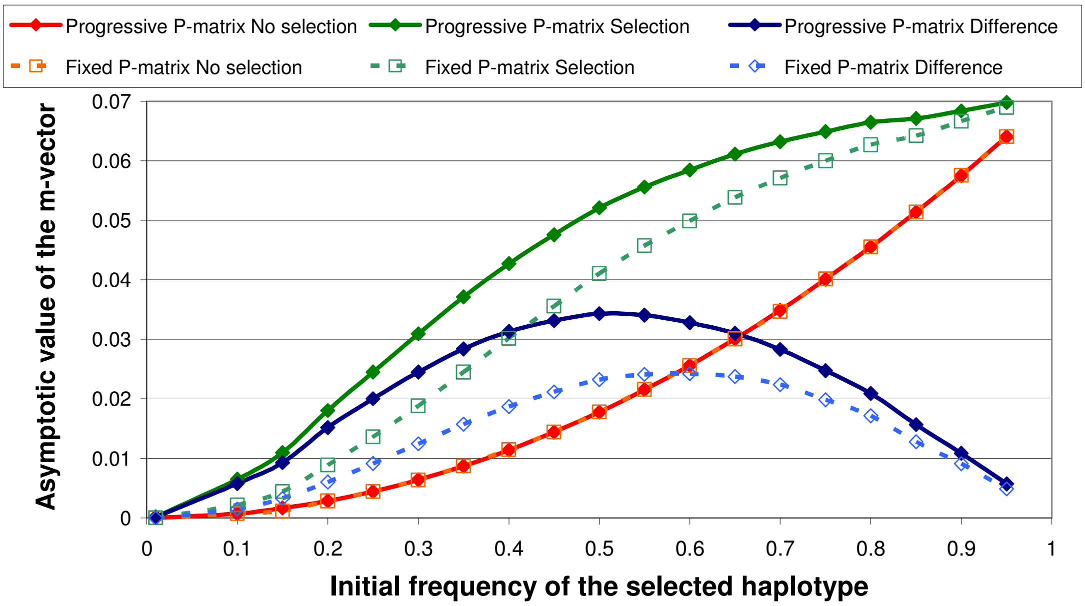

Figure 2: Asymptotic value of the gene-proportion depending on the initial haplotype frequency and the selected proportions with progressive or fixed P-matrices.

The results for the SDE are presented in Figure 3. It appeared that the higher the initial haplotype frequency, the better the final SDE value. Moreover, the selection in the breeding nucleus clearly increased the SDE-value achieved at the time horizon. The difference was highest for intermediate haplotype frequencies (0.19 for an initial frequency of 0.5). The increase due to selection relatively to an unselected breeding nucleus achieved 335% for an initial haplotype frequency of 0.1. However, the difference between a selected and an unselected breeding nucleus was smaller for high initial haplotype frequencies.

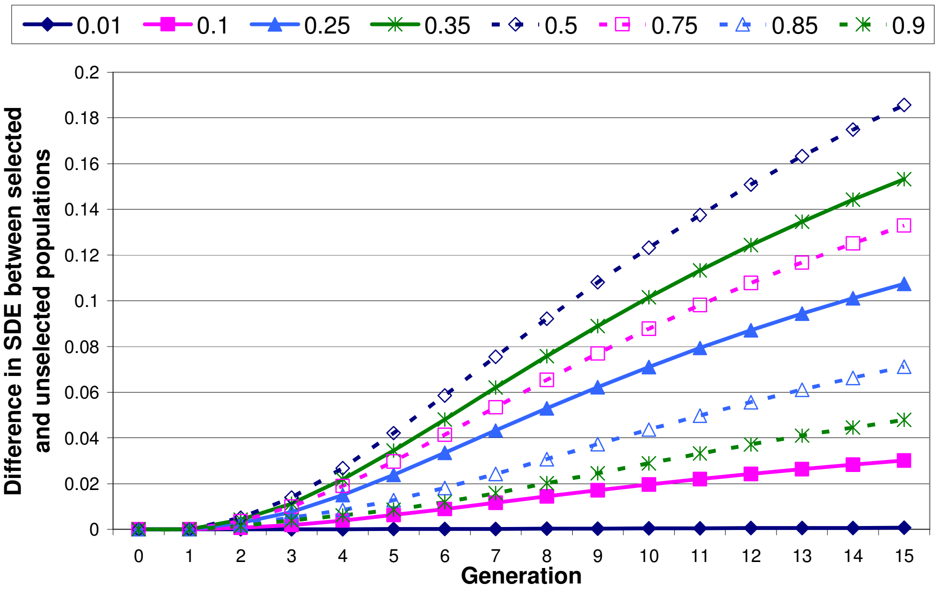

Figure 3: Differences between the SDE-values of unselected and selected breeding nuclei over the generations for various initial frequencies of the selected haplotype

DISCUSSION

We investigated the impact on the changes in gene frequencies due to selection on the SDE and gene proportions achieved in a breeding program. We first could establish that using a fixed P-matrix leads to an underestimated SDE and gene proportions when the genotype frequencies are changing in the breeding nucleus, which may induce false decisions when deciding how to conduct a breeding program. It thus seems necessary to account for these changes in the gene-flow method. As expected, there was little difference between in the unselected situation between using a progressive or a fixed P-matrix, meaning that both methods can be indifferently used if the genotype frequencies remain unchanged in the breeding nucleus. This could happen in the case of a selection based on traits independent from the genes studied.

It also appeared that the initial allele haplotype frequency highly influenced the results which were better for high frequencies. The expected gain in selection is higher for intermediate frequencies. On one hand, a plateau seems to be reached when starting with a high haplotype frequency. On the other hand a certain number of generations may be required to have a real increase of the haplotype frequency in the production population if the selected haplotype is initially rare and thus to have an obvious increase in the breeding values due to the haplotype. Even in cases where the achieved SDE or asymptotic gene proportion remained low, the value achieved may be more than the double (SDE) or 4 times (m-vector) the one observed with an unselected breeding nucleus.

It thus seems necessary to precisely describe the gene frequencies in the breeding nucleus when starting a breeding program including them in the breeding goal. Further investigation must be conducted to evaluate the influence of the selection intensity in the breeding nucleus and the genetic progress achieved in the breeding programs when using a progressive P-matrix in the gene-flow method.

CONCLUSION

It results from this study that the actual knowledge and inclusion of genes in the breeding objectives should lead to changes in the gene-flow methodology classically used to evaluate a breeding program. In order to objectively compare possible breeding programs, one should know the initial haplotype frequencies in the breeding nucleus and include their estimated evolution in the P-matrix.

ACKNOWLEDGMENTS

This study was conducted as part of the FUGATO+brain project and was financially supported by the German ministry for education and research (BMBF), Förderverein Biotechnologieforschung e.V., (FBF Bonn), Lohmann Tierzucht GmbH (Cuxhaven), Verband hannoverscher Warmblutzüchter e.V., Trakehner Verband, Verband des Oldenburger Pferdes e.V., Verband der Züchter des Holsteiner Pferdes e.V. and Westfälische Pferdestammbuch e.V..

REFERENCES

Börner V., Reinsch N.: **Gametic gene flow method accounts for genomic imprinting.** J. Anim. Breed. Genet. 2009, published ahead of print, 1-15.

Chakraborty R., Moreau L., Dekkers J.C.M.: **A method to optimize selection on multiple identified quantitative trait loci**. Genet. Sel. Evol. 2002, 34: 145-170.

Elsen J.M., Mocquot, J.C. **Méthode de prévision de l'évolution du niveau génétique d'une population soumise à une opération de sélection et dont les générations se chevauchent**. Bull. tech. Dépt. Génét. Anim., INRA, 1974, 17: 30-54

Hill W.: **Prediction and evaluation of response to selection with overlapping generations.** Anim. Prod. 1974, 18: 117-139.

Leymarie C., Astruc J.M., barillet F., Bibe B., Bonnot A., Bouffartigue B., Bouix J., Dion F., Francois D., Jouhet E., Jullien E., Moreno C., Bosher M.Y., Palhiere I., Raoul J., Tiphine L., Bouchel D., Chibon J., Raynal A., Tribon P.: **The French national breeding plan for scrapie resistance: an example of molecular information used in selection.** 60th annual meeting of the European Association for Animal Production, Barcelona, 24-27 August 2009.

Ytournel F., Simianer H.: **Deterministic and stochastic prediction of the evolution of QTL genotype frequencies with overlapping generations and finite population size**. 60th annual meeting of the European Association for Animal Production, Barcelona, 24-27 August 2009.