59

Original article

On the relevance of three genetic models for the description of genetic variance in small populations undergoing selection

Florence Fournet-Hanocq [*], Jean-Michel Elsen

Station d'amélioration génétique des animaux, Institut national
de la recherche agronomique, 31326 Castanet-Tolosan cedex, France

Abstract – The conservation of genetic variability is recognized as a necessary objective for the optimization of selection schemes, particularly when populations are small. Numerous models, differing by the genetic model they rely on, are available to better understand and predict the evolution of genetic variance in a small population undergoing selection. This paper compares three genetic models, treated either analytically or with Monte-Carlo simulations, first in order to validate the predictions provided by a 'full-finite model' for well-known phenomena (e.g. the effect of population management on genetic variability), and second, to evaluate when and how the assumptions made in the two analytical models induce the departure from the third model. The FFM is shown, first, to be in close agreement with the Gaussian theory when used with a large number of loci, the stochastic approach making it much more flexible than the two algebraic models. In the second part of the study, the infinitesimal model appears to be more robust than the semi-infinitesimal one. Major sources of discrepancy between the deterministic models and the FFM are identified, notably the hypothesis of independence between loci, and then the infinite number of loci or alleles per locus. © Inra/Elsevier, Paris

modelling / genetic variance / selection / small population

Résumé – De l'intérêt de trois modèles génétiques pour la description de la variance génétique dans les petites populations sélectionnées. La conservation de la variabilité génétique est reconnue comme un objectif nécessaire pour l'optimisation des schémas de sélection, notamment pour les petites populations. De nombreux modèles, différant par les hypothèses de déterminisme génétique sur lesquels ils reposent, sont disponibles pour une meilleure compréhension et une meilleure prédiction de l'évolution de la variabilité génétique dans une petite population soumise à sélection. Cet article compare trois modèles génétiques différents, traités soit par la voie analytique soit par des simulations Monte-Carlo, d'une part pour valider les prédictions fournies par un «modèle fini complet» pour des phénomènes connus (comme l'influence de la gestion de la population sur la variabilité génétique), d'autre part pour évaluer quand et comment les hypothèses faites

---

[*] Correspondence and reprints

dans les deux modèles analytiques induisent un écart avec le dernier modèle. Le modèle fini complet apparaît, dans un premier temps, en bon accord avec la théorie gaussienne quand il est utilisé avec un grand nombre de locus, l'approche stochastique le rendant de plus beaucoup plus souple que les deux modèles algébriques. Dans la seconde partie de l'étude, le modèle infinitésimal apparaît plus robuste que le modèle semi-infinitésimal. Des sources majeures d'écart entre les modèles déterministes et le modèle fini complet sont identifiées, notamment l'hypothèse de l'indépendance entre les locus, puis le nombre infini de locus et d'allèles par locus. © Inra/Elsevier, Paris

modélisation / variance génétique / sélection / petite population

1. INTRODUCTION

Genetic variability is necessary to provide genetic progress through selection and the conservation of genetic variability is of increasing concern in the optimization of selection schemes, particularly when selected populations are small. Numerous experiments [1, 6] have shown a decrease in genetic variability over time, due to genetic drift and selection, until the exhaustion of variance in many cases. Therefore, thorough knowledge of the evolution of genetic variance in relation to these two phenomena is important when constructing optimal selection schemes.

Several kinds of models are available to describe the evolution of genetic variability, depending on the hypotheses used concerning genetic determinism. Analytical models, whose properties often rely on the hypothesis of normality, provide a quite simple formalization of the phenomena acting on genetic variance. Monte-Carlo simulations allow a more detailed and complex description of polygenic inheritance.

This paper aims to compare the predictions of genetic variability in a small population undergoing long-term selection provided by three different genetic models, two of them treated analytically and the last with Monte-Carlo simulations. This comparison concerns essentially their sensitivity to the characteristics of the genome. The genetic models underlying the Monte-Carlo simulations and the two analytical models are presented first. The predictions provided by the three models will then be compared first to validate the genetic model used in the MC simulation when it approaches the infinitesimal hypothesis, and secondly to determine when the analytical models depart from the 'full-finite model', used here as reference.

2. DESCRIPTION OF THE MODELS

Three genetic models are presented, two of them being treated analytically and the last with Monte-Carlo simulations. The Monte-Carlo model will be developed first as the most complete situation, where the effects of genetic drift and selection are not algebraically formalized but implicitly accounted for in the simulation. The analytical models of Verrier et al. [14] and Chevalet [3] will then be described, each corresponding to a different choice of hypotheses for the algebraic formalization of the phenomena acting on variance.

2.1. The 'full-finite model' (or FFM)

The genetic model underlying the simulation is based on the assumption that a selected trait is controlled by a finite number of linked genes individually identified, located on chromosomes, with a finite number of alleles per locus. This model is then called the 'full-finite model' or FFM. Individual genotypes are generated according to the number of loci, alleles per locus, loci per chromosome, recombination rates and relative effects of the loci on the selected trait. For a given individual, an environmental value (assumed to be normally distributed) is added to its generated genotype, giving its phenotypic value. A within-generation mass selection is based on these phenotypes. Selected breeding individuals are randomly mated and produce a new generation. This is performed through the simulation of meiosis and pairing of gametes. Generations are assumed to be discrete.

The simulation algorithm, initiated by Hospital [8] and developed by Fournet et al. [7], uses the Monte-Carlo principle and provides the mean values and standard deviations for genetic mean and variance over time.

2.2. Analytical models

The first model presented, established by Verrier et al. [14], relies on the hypothesis that the selected trait is controlled by an infinite number of independent loci, with identical and small effects. It will then be called in the following 'infinitesimal model' or IM. The computation of the inbreeding coefficient in this model accounts for the effect of selection on family structure (for more details, see [14]).

The second model, developed by Chevalet [3] for a monoecious population of N individuals, assumes a finite number of unlinked loci L with an infinite number of alleles per locus, and will be called the 'semi-finite' model or SFM. The joint distribution of the gene effects is then assumed to be multivariate normal. Derivations of the variances of gene effects and covariances between gene effects under selection lead to the prediction of the joint evolutions of genetic and genic variances by two recurrence equations. The number N of individuals has been replaced here by the effective size of population N_e, derived from the Latter-Hill equation [11], with a Poisson distribution of the number of offspring for each genealogical path, accounting for genetic drift.

3. CASES STUDIED

A polygenic-like situation with an infinite number of alleles per locus was first simulated in the FFM, to check if the evolution of genetic variance given by the three models was similar. This basic system was defined as a population of size N (with as many males as females), evaluated on their own performances for each generation, with, respectively, 25 and 50 % males and females retained. The heritability of the selected trait was assumed to be 0.3. A thousand independent loci with 500 alleles per locus were simulated in the Monte-Carlo model. The sensitivity of the models to deviations from the basic situation was then evaluated. The comparison criterion was the ratio between genetic variances at generation t and generation 0 ($RV^{[t]}$).

3.1. Population size

Different sizes of candidate population N (total sizes of 96, 192 and 480, with as many males as females for each size) were tested.

3.2. Parameters of genetic determinism

In this study 1 000, 100, 10 and 2 independent loci were assumed in the FFM and the SFM. This comparison of sensitivity to the number of loci was performed assuming that all loci were located on one chromosome in the FFM.

Here 100, 10 and 2 alleles per locus, in the case of 100 or 10 loci, were simulated in the FFM in order to show the deviation of the SFM from the FFM when this parameter decreases.

The recombination rate r was assumed to be 0.5, 0.1, 0.01 and 0.001, in the case of 1 000, 100 and 10 multiallelic loci controlling the selected trait, in order to check the effect of linkage on the prediction of genetic variance over time in the FFM. The effect of linkage was not studied in the model of Chevalet, as assuming $r_{ij} \neq 0.5$ for any pair of loci (i, j) would produce as many equations as different pairs of loci.

In the 'full finite' model, the relative contributions of the loci to the genetic variance of the selected trait were assumed to be identical in the preceding comparisons. To test the robustness of the results with respect to this assumption, and following Lande and Thompson (1990), variances were assumed to follow a geometric series, with the lth locus contributing $VA(1-a)a^{l-1}$, where the constant a determines the relative magnitude of the contributions of each locus. This constant is related to the effective number of loci as: $L_e = (1+a)/(1-a)$. Simulations were performed with 1 000 biallelic loci, with effects following a geometric series where the parameter a was given values corresponding to 2, 5, 10, 50 and 100 effective loci. The resulting evolutions of genetic variability were compared to the corresponding curves of loci with identical contributions.

Thirty generations of selection were simulated. A hundred simulations were performed for each combination of factors. With as many simulations, a difference just higher than 5 % between the predictions of the different models would be significant (at the 5 % level). A 10 % difference was then considered as significantly larger.

4. RESULTS

The predictions provided by the three models for the joint effects of population size and selection intensity on the evolution of genetic variance were compared. *Figure 1* illustrates the case of intermediate value of selection intensity (i.e. 25 % of males selected), as the trend was the same for the other two tested values (50 % and 6.25 % of males selected).

It can be pointed out that the three models provided almost the same evolution of genetic variance, whatever the population size and selection intensity. As expected [4], the higher the selection intensity, the greater the influence of the population size, and the higher the decrease in genetic variance. This comparison of the
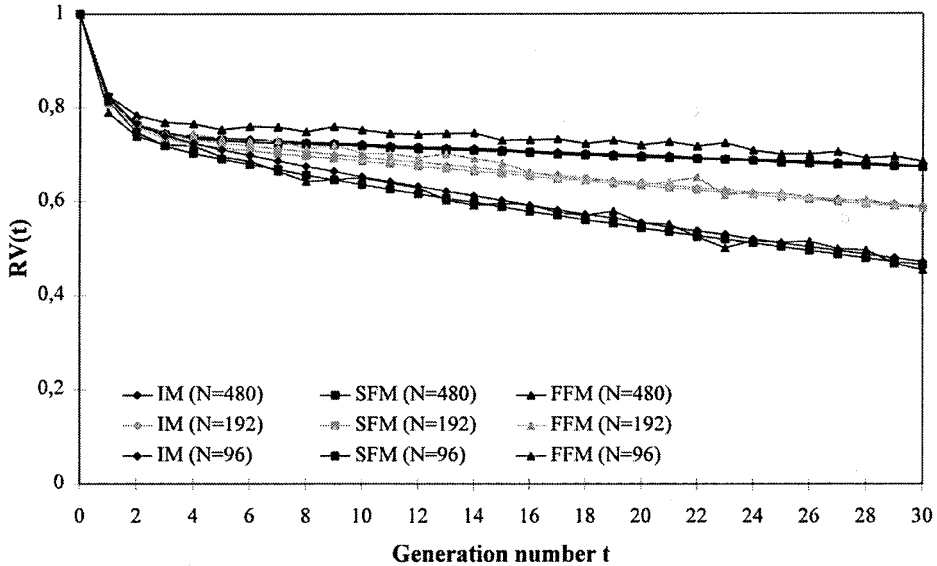
Figure 1. Evolution of genetic variance with population size in three models; IM for 'infinitesimal', SFM for 'semi-finite' and FFM for 'full-finite' models; selection intensity is constant and the number of loci is very large in SFM and FFM.

predictions given by analytical and stochastic models for well-known phenomena allowed validation of the genetic model used for the MC simulations.

4.1. Effect of number of loci

Table I presents the values of $RV^{[30]}$ provided by the three models when decreasing the number of loci L (with multiallelic loci in the FFM). $RV^{[30]}$ in the IM departed by more than 10 % from the FFM prediction for studied values of L lower than 250, while the predictions of SFM and FFM were in quite good agreement for more than 50 loci. Although the semi-infinitesimal model accounts for a finite number of loci, its sensitivity to this parameter is quite weak: the difference between IM and SFM exceeds 10 % only when the number of loci considered in the SFM is lower than 10. This behaviour must be related to the hypothesis of infinite number of alleles per locus. Results of FFM presented later, where the effect of the number of loci decreases when increasing the number of alleles per locus, are consistent with this observation. *Figure 2* illustrates the evolution of $RV^{[t]}$ over time for 1 000, 100, 10 or 2 loci. For 100 loci, IM and SFM depart from FFM only after 20 generations, while for 10 loci, they both depart from FFM as early as generation 7 and with 2 loci, IM and SFM depart from FFM after generations 3 and 6, respectively. The additive genetic variance in FFM decreased dramatically with a very small number of loci (10 and 2). This result was of course expected and it illustrates the strong influence of the infinitesimal hypothesis on the predictions.

Table I. Values of the additive genetic variance at generation 30, expressed in percentage of the initial additive genetic variance ($RV^{[30]}$), for different numbers of loci L in the three models.

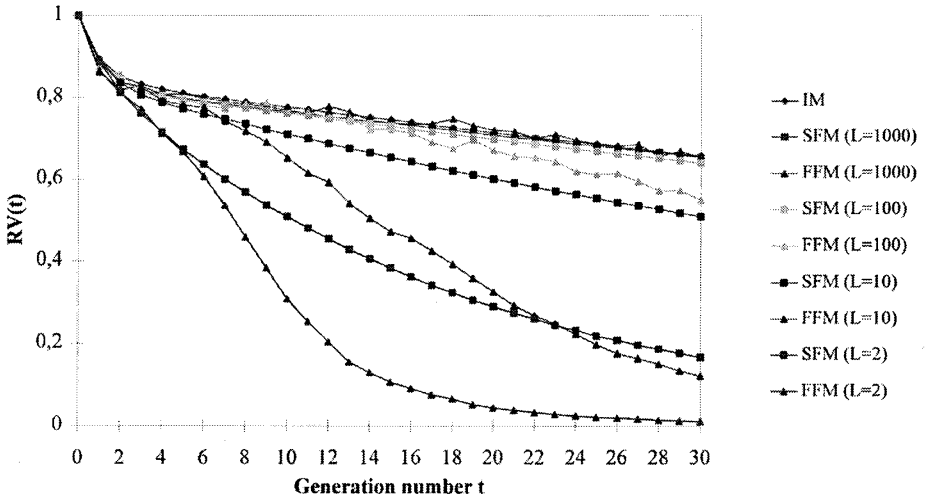| L	1 000	500	250	150	100	50	10	5	2
$RV^{[30]}$ for the 'infinitesimal' model	65.6 %	65.6 %	65.6 %	65.6 %	65.6 %	65.6 %	65.6 %	65.6 %	65.6 %
$RV^{[30]}$ for the 'semi-finite' model	65.2 %	65.2 %	64.8 %	64.4 %	63.8 %	62.2 %	50.8 %	39.0 %	16.6 %
$RV^{[30]}$ for the 'full-finite' model	65.0 %	58.8 %	55.4 %	53.3 %	54.8 %	45.1 %	12.1 %	5.1 %	1.1 %

Figure 2. Evolution of genetic variance with the number of multiallelic loci in three models; IM for 'infinitesimal', SFM for 'semi-finite' and FFM for 'full-finite' models; $N = 192$, $A = 500$ alleles/locus in the FFM.

However, the infinitesimal model remains quite robust so long as the number of loci assumed is not too small.

4.2. Effect of the number of alleles per locus

Figure 3a, b indicates a loss of genetic variance in the FFM, when the number A of alleles decreased, higher when the number of loci was smaller: the genetic variance decreased for fewer than 10 alleles per locus for $L = 100$ and
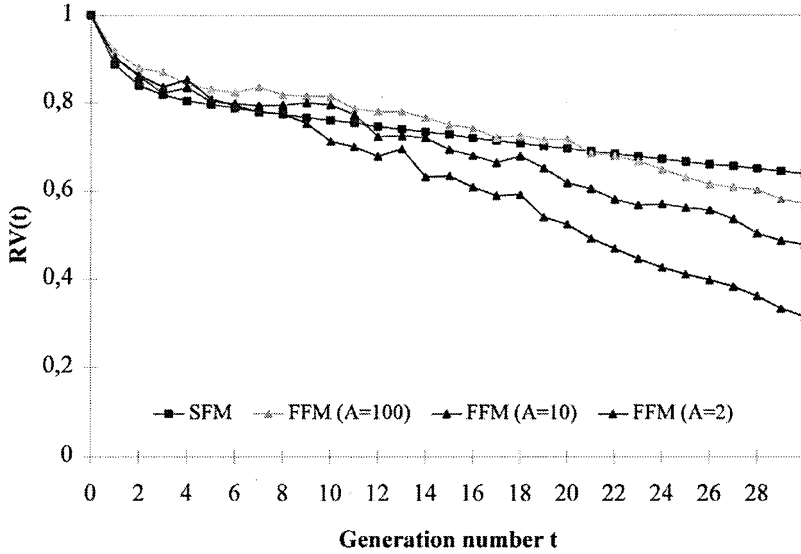
Figure 3a. Evolution of genetic variance for 100 loci and three numbers of alleles in two models; SFM for 'semi-finite' and FFM for 'full-finite' models; $N = 192$.
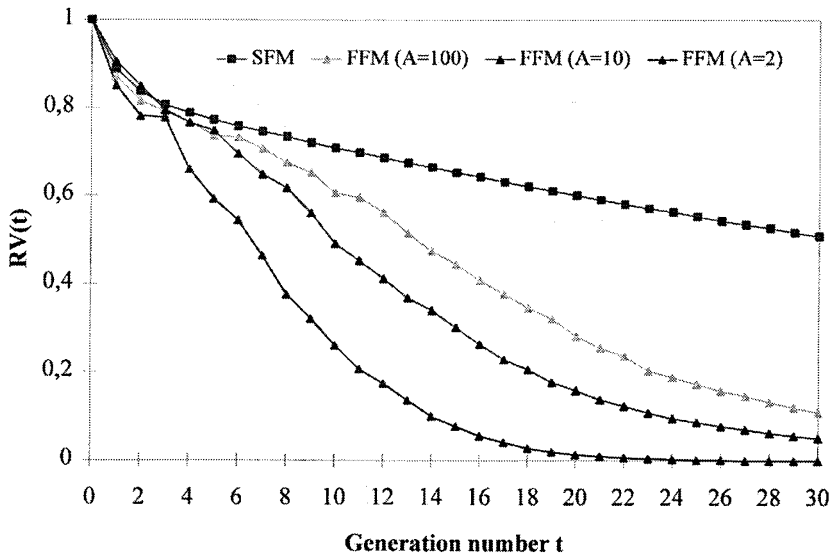

Figure 3b. Evolution of genetic variance for 10 loci and three numbers of alleles in two models; SFM for 'semi-finite' and FFM for 'full-finite' models; $N = 192$.

since 100 alleles per locus for $L = 10$. A combination of the two variables indicating the amount of available variability influenced the behaviour of the genetic variance over time, more than did each variable alone: at a given generation, 1 000 biallelic loci and 100 loci with 50 alleles per locus provided the same $RV^{[t]}$ (results not shown).

4.3. Effect of linkage between loci

The decrease in genetic variance when increasing the linkage between loci was higher when the number of loci was lower. For 1 000 multiallelic loci, only very strong linkage ($r = 0.001$) gave a significant decrease in the predicted $RV^{[30]}$ in the FFM (38 % lower than for $r = 0.5$). For 100 (*figure 4*) or 10 multiallelic loci, no difference was observed between the absence of linkage and a recombination rate of 0.1, but below $r = 0.01$, stronger linkage led to a higher decrease in genetic variance in the early generations and a lower asymptotic plateau in the latter ones. In fact, a large number of linked loci behave as a small number of independent loci. To a certain extent, the genome size seemed the main factor acting on genetic variance, in agreement with Robertson [12] who showed that the limit in genetic response for a given recombination rate was directly linked to the chromosome length. Nevertheless, the comparison remained limited as Robertson inferred a general description of the selection process in finite populations only for a large number of loci.
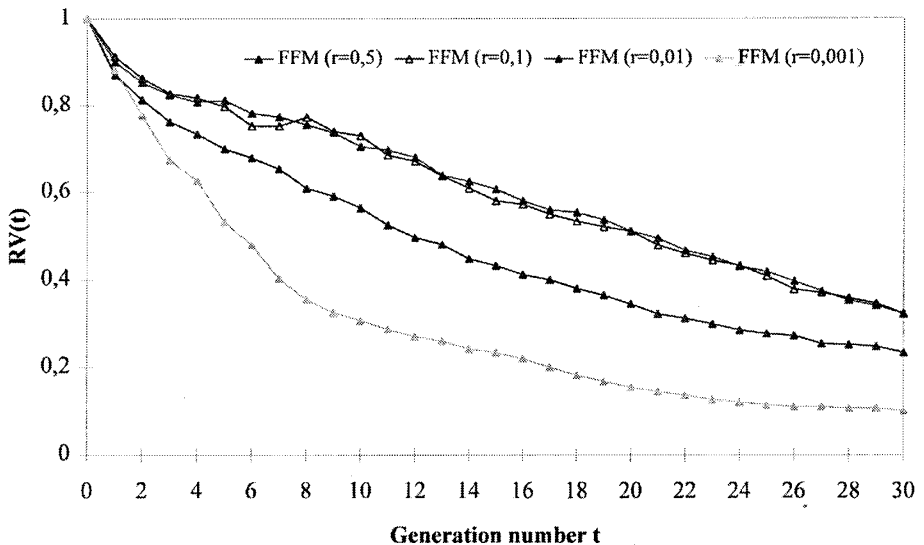

Figure 4. Evolution of genetic variance with recombination rate for 100 loci in the 'full-finite' model; $N = 192$, $A = 500$ alleles/locus in the FFM.

4.4. Effect of inequality between loci effects

Predictions of the genetic variance for 1 000 biallelic loci whose effects followed a geometric series, with varying parameter a giving a number L_e of effective loci, and predictions for $L = L_e$ loci with identical effects are compared in *figure 5*. It can be seen that the predictions for $L = x$ loci of identical effects and for 1 000 loci of differential effects giving $L_e = x$ effective loci were very close, whatever the value of x. This showed that even 1 000 loci could not be considered an infinitesimal genome, if the hypothesis of differential effects of the loci (a few loci with relatively large effects and many others with small effects) is true [10].
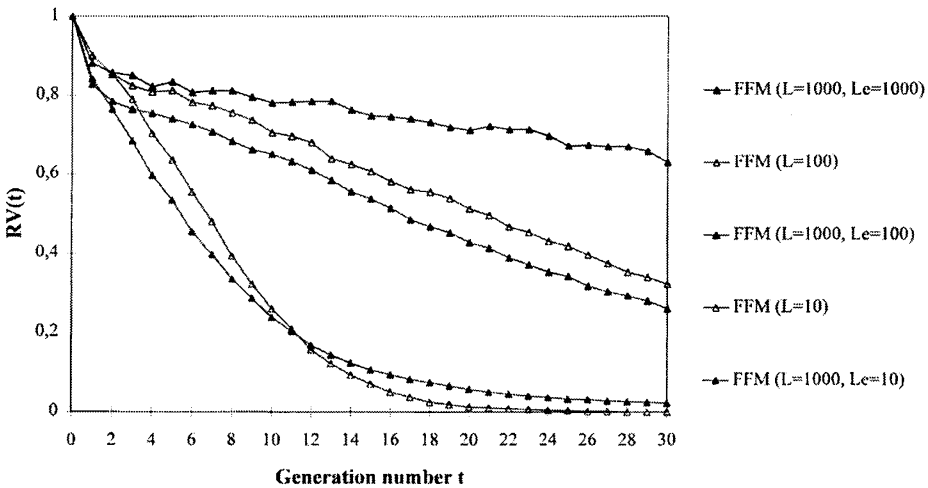

Figure 5. Evolution of genetic variance for 1 000 biallelic loci with the distribution of the effects of loci in the 'full-finite' model; $N = 192$.

5. DISCUSSION AND CONCLUSION

The first part aimed to investigate the validity of the genetic model underlying the MC simulations, when compared to analytical models corresponding to different hypotheses on the genetic determinism. The first result to be pointed out was that the three models, when a large number of loci were assumed for the 'semi-finite' and 'the full-finite' models, were in close agreement, whatever the population size and selection intensity assumed. Moreover, the computation of the effective population size for the monoecious model of Chevalet seemed to be quite valid as it provided the same prediction as the dioecious models. This part verified that under the hypothesis of a very large number of independent identical loci, the oligogenic model and the Gaussian theory agree closely. These first results validate the use of the 'full-finite' model in infinitesimal conditions, as its predictions seem reliable and the stochastic approach makes it much more flexible than the other two models.

Furthermore, the stochastic approach intrinsically takes account of the reduction in selection intensity as compared with the theory, of the relationships between mates and inbreeding induced in the offspring and of the changing variance of gene effects due to genetic drift and faster or slower fixation of alleles. The need of unknown parameters to introduce in the model does not prevent its use, as the other two models also make assumptions on the unknown parameters they use. And the increasing knowledge on QTLs implied in the genetic determinism of selected traits will provide elements for such a modelling.

In the second part, the hypothesis of an infinite number of loci controlling the selected trait was shown to be very important: similar prediction of $RV^{[t]}$ with infinitesimal and 'full-finite' models was obtained only when simulating more than 500 loci in the FFM. Under this value, the departure of the Verrier et al. [14] model from the FFM was substantial. One can wonder if such a large number of loci (500 or 1 000) controlling one trait is realistic? Tanksley [13] suggests that the number of QTL implicated in various traits for a large set of animal or vegetable species should vary between 1 and 18. But famous selection experiments on maize [5] expressed genetic progress over 76 generations, suggesting that genetic variability is infinite according to the infinitesimal theory. In animal breeding also, selection experiments have been based on this theory for a long time and no evidence of important discrepancy between results and theory has been found. This contradiction strengthens the value of estimating the effective number of QTL controlling a trait. It also points out the lack of complexity in the FFM hypotheses: adding interactions between genes or mutations, or locating the genes on several chromosomes, might be more realistic and informative. Indeed, recent studies [2, 9] investigated the contribution of mutations as a source of new genetic variance in mice. But, in our study, since the other two models did not account for interactions or for new genetic variance arising from mutation, this hypothesis was excluded in this version of FFM.

The second hypothesis of an infinite number of alleles per locus may explain the discrepancy observed between the 'semi-finite' and the 'full-finite' models for various numbers of loci. The total amount of available genetic variability, i.e. the combination between the numbers of loci and alleles per locus, is not taken into account in the same way in the two models. Is an infinite number of alleles per locus more realistic than an infinite number of loci? Mutations might lead to a huge number of very close allelic forms, whose effects are normally distributed; on the contrary, the large potential variability is reduced because some of the mutational allelic forms are not functional, and different alleles give exactly the same phenotype. In the end, a discrete distribution of allelic effects is possible. In this case, the prediction provided by the 'semi-finite' model holds true only in the short term, with a very slow decrease in variance after the first generations of selection, whereas the 'full-finite' model shows a clear tendency to a rapid fixation of available alleles.

The effect of linkage was also studied by Hospital [8]. Two phases can be observed in our results when the linkage is strong: a first phase with a rapid decrease in genetic variability and a second phase of slow decrease reaching a plateau. Hospital [8] explained this phenomenon by a weak rearrangement of gametes by recombination. Selection first sorts the most favourable combinations in the initial

genetic pool, with a rapid reduction in genetic variability, and produces hitch-hiking phenomena: unfavourable genes are selected jointly with favourables ones and are poorly eliminated because of the low recombination rate in the second phase. When linkage is less intense, the gametes are reorganized by recombination over a longer period and the decrease in genetic variance is more regular. This part of the study demonstrates a major source of possible discrepancy from reality when using the two analytical models: as they both consider independent loci, they are likely to overestimate the remaining genetic variance over time. It may then be underlined that the ability to consider a possible linkage between loci confers a greater concern to the genetic model in the FFM, as linkage is known to exist between loci and is notably an essential principle for QTL detection. Finally, the identical effects of the loci appeared to be a strong hypothesis. Indeed, following Lande and Thompson [10], the distribution of gene effects according to a geometric series pointed out that the number of effective loci was of greater concern than the real number of loci. This point is also of great concern in this study. Indeed, results of molecular genetics indicate, for most of the quantitative traits under study, the influence of a mixed heredity, i.e. a small number of genes with large effects and a large number of genes with small effects. But this kind of genome structure is not easy to integrate in analytical models. The approach developed by Chevalet [3] accounts for differential distribution of gene effects, by joining genes with variable contributions to the genetic variance of the trait into several independent clusters of genes with equal contributions. Moreover, within one given group of genes in this model, the loci may be closely linked. The main disadvantage of this approach is the overestimation of genetic variance after the first generation, probably due to the departure of the gene effects from Gaussian distribution. By accounting for a finite number of loci and a finite number of alleles per locus, the 'full-finite' genetic model developed in this paper, with a stochastic approach, seems to be an easy and quite consistent way for studying the behaviour of genetic mean and variance in a situation of mixed heredity.

REFERENCES

[1] Al Murrani W.K., Roberts R.C., Genetic variation in a line of mice selected to its limit for high body weight, Anim. Prod. 19 (1974) 273–289.

[2] Caballero A., Keightley P.D., Hill W.G., Accumulation of mutations affecting body weight in inbred mouse lines, Genet. Res. 65 (1995) 145–149.

[3] Chevalet C., An approximate theory of selection assuming a finite number of quantitative trait loci, Genet. Sel. Evol. 26 (1994) 379–400.

[4] Crow J.F., Kimura M., An Introduction to Population Genetics Theory, Harper and Row, New York, 1970.

[5] Dudley J.W., Seventy-six generations of selection for oil and protein percentage in maize, in: Pollack E., Kempthorne O., Bailey T.B. Jr (Eds.), Proceedings of International Conference on Quantitative Genetics, Iowa State Univ. Press, Ames, 1977, pp. 459–473.

[6] Falconer D.S., Improvement of litter size in a strain of mice at a selection limit, Genet. Res. 12 (1971) 215–235.

[7] Fournet F., Hospital F., Elsen J.M., A FORTRAN program to simulate the evolution of genetic variability in a small population, Comput. Applic. Biosci. 11 (1995) 469–475.

[8] Hospital F., Effets de la liaison génique et des effectifs finis sur la variabilité des caractères quantitatifs sous sélection, thèse de Doctorat, Université Montpellier II.

[9] Keightley P.D., Hill W.G., Quantitative genetic variation in body size of mice from new mutations, Genetics 131 (1992) 693–700.

[10] Lande R., Thompson R., Efficiency of marker-assisted selection in the improvement of quantitative traits, Genetics 124 (1990) 743–756.

[11] Ollivier L., Le calcul de l'effectif génétique des populations animales, Ann. Génét. Sél. Anim. 5 (1973) 363–368.

[12] Robertson A., A theory of limits in artificial selection with many linked loci, in: Kojima K. (Ed.), Mathematical Topics in Population Genetics, Springer-Verlag, Berlin, 1970, pp. 246–288.

[13] Tanksley S.D., Mapping polygenes, Annu. Rev. Genet. 27 (1993) 205–233.

[14] Verrier E., Colleau J.J., Foulley J.L., Predicting cumulated response to directional selection in finite panmictic populations, Theor. Appl. Genet. 79 (1990) 833–840.