Analysis of the within-population genetic diversity and the effective size based on different kinds of information in selected animal populations

Valérie Loywyck

UFR Génétique,
Elevage et Reproduction

Agriculture, Alimentation
Biologie, Environnements, Santé

UMR Génétique
et Diversité Animales

Thesis

to obtain the degree

DOCTEUR D'AGROPARISTECH

Field: Animal Genetics

presented and defended by

Valérie LOYWYCK

December, 18, 2007

Approches multicritères de l'évolution de la variabilité génétique et de la taille efficace au sein des populations animales soumises à sélection

Analysis of the within-population genetic diversity and the effective size based on different kinds of information in selected animal populations

Supervisor: Etienne VERRIER

with the co-supervison of
Piter BIJMA (WUR), Marie-Hélène PINARD-van der LAAN (INRA), Jacques DAVID (INRA)

Jury

Miguel Angel TORO	Senior scientist, INIA, Madrid	Referee
Philippe BARET	Professor, Université Catholique de Louvain	Referee
Joël CUGUEN	Professor, Université de Lille	Examiner
Mathilde DUPONT-NIVET	Scientist, INRA, Jouy-en-Josas	Examiner
Piter BIJMA	Associate Professor, Wageningen University	Examiner
Etienne VERRIER	Professor, AgroParisTech	Examiner

Remerciements

Je remercie les membres du jury : Miguel Toro et Philippe Barret qui m'ont fait l'honneur d'accepter d'être rapporteurs, ainsi que Mathilde Dupont-Nivet et Joël Cuguen, examinateurs de ce travail.

Je remercie celui qui fût à la fois mon directeur de thèse, mon responsable d'enseignement et directeur d'UMR : Etienne Verrier.

Je suis aussi reconnaissante aux différentes unités qui ont eu la gentillesse de m'accueillir tout au long de ma thèse :

- merci à l'UMR « Génétique et Diversité Animales », à Jouy ;
- merci à l'UMR « Diversité des Plantes Cultivées », à Montpellier ;
- thanks to the « Animal Breeding and Genomics » group in Wageningen.

Cette thèse « voyageuse » a été rendue possible par le financement de la Direction Scientifique de l'INA P-G, le financement de la Mission des Relations Internationales de l'INRA et un contrat BRG.

Merci à Jacques David de m'avoir ouvert les portes de la station de Mauguio et d'avoir accepté de faire partie de mon comité de thèse, malgré un emploi du temps bien chargé. Thanks to Johan van Arendonk for letting me part for a while of the ABG group at Zodiac.

Je tiens également à remercier Marie-Hélène Pinard-van der Laan, Bertrand Bed'Hom, et Marie-Hélène Müller pour leur soutien, aussi bien amical que professionnel, et leurs conseils au quotidien. Thanks to Piter Bijma for his daily and warm support and his helpful advice.

Merci également à Isabelle Goldringer qui faisait partie de mon comité de thèse et avec qui j'ai fait mes premiers pas dans le monde (pas si éloigné que ça) de la génétique des plantes.

Et puis il y a tous ceux des différentes unités (à Montpellier et à Jouy mais aussi à l'Agro) que je n'ai pas cités mais qui ont rendu conviviaux et enrichissants mes séjours parmi

eux…Merci à tous! Thanks to all Wageningen people who made my stays as friendly and interesting as possible!

Je voudrais aussi remercier Fabienne Dedole-Porcherot pour m'avoir aidé dans l'imbroglio des formalités administratives et de déplacement, Agathe Vieaud pour sa bonne humeur et nos discussions « de filles » dans le bureau, et Thierry Heirmann pour ses astuces informatiques et automobiles !

Enfin, il y a ceux qui m'ont toujours soutenue : mes parents, toujours compréhensifs et présents, même par téléphone ou par internet quand j'étais loin! Heureusement qu'il y avait Free !!

Merci à tous mes amis, de Paris et des quatre coins de la France : merci d'être toujours là, même si je n'étais pas toujours des plus disponibles… J'ai une pensée toute particulière pour ma chère Hortense, toujours optimiste et attentive.

Et puis Jérôme… merci pour ton amour, ta patience et ton soutien indéfectible.

Merci !!!

Thanks !!!

Abstract

The scientific purpose of the thesis is twofold: (*i*) to investigate in details the links between criteria originating from different kinds of information, such as molecular markers, pedigrees, or phenotypes for quantitative traits, (*ii*) to go further in the examination of the joint evolution of the neutral variability and the selected variability. According to the kind of information, genetic diversity and its evolution are described through various parameters dealing with more or less complex underlying model, thus, we will have more or less realistic representation of the genetic diversity. On one hand, analysis of polymorphism gives a direct overview of the diversity: genotypes and allele frequencies of candidate gene or molecular markers will give access to specific polymorphism of known areas in the genome, whereas pedigrees will give access to anonymous polymorphism of neutral unknown areas in the genome. On the other hand, analysis of phenotypes gives a general overview of genetic diversity, assuming the model of representation to be more complex.

The principal methods that are available for the analysis of genetic diversity and those used along this thesis, using different kinds of information are presented in chapter 1.

Chapter 2, 3 and 4 deal with a valorisation of different kinds of information from experimental chicken lines, selected for immune response traits, for the integrate analysis of genetic variability.

Genetic gain and genetic diversity based on performances and pedigree data were analysed in chapter 2. The observed evolutions of inbreeding and genetic gain were compared to values predicted by some theoretical models. The deterministic methods compared yielded results that were close to those observed in real data and differences between theoretical predictions and experimental results mainly arise from differences between the true and the assumed selection scheme, and from mathematical simplifications applied in the prediction methods. Effect of selection scheme on inbreeding and other criteria of genetic variability, based on pedigree data, was also investigated. The effective number of ancestors appeared to be the most relevant parameter in monitoring genetic diversity using pedigree information since it takes into account the loss of genetic diversity due to genetic drift occurring during the pedigree development.

Estimation and evolution of genetic parameters and polymorphism evolution of a candidate gene were handled in chapter 3. Confrontation of different theoretical predictions with

observed evolution of the polymorphism within lines and analysis of variance coponents were undergone to check the neutrality of MHC for the traits in our selected populations. But beyond the interest of MHC effect in immune responsiveness, this study has highlighted the interest of combining various approaches to assess the effect of a candidate gene and the evolution of its polymorphism, especially in the case of rare alleles. Changes in additive genetic variance during the course of the experiment were also investigated and we examined the cope of the underlying model (polygenic infinitesimal model) used for the estimation of the genetic parameters. Increasing the number of generation or taking subsets of generations using REML may be an appropriate method in monitoring genetic diversity over time and infer about the effect of selection on reduction of additive genetic variance.

Evolution of polymorphism of supposedly neutral or selected molecular markers was analysed and compared in chapter 4. Different methods were combined, statistical analysis as well as modelling and simulations, to detect signature of selection left by QTL. Pictures of genetic diversity were drawn from polymorphism evolution of markers located in QTL regions and supposedly neutral markers, which may be considered as a reference. This study has shown that QTLs are very sensitive to the trait they related to and that a marker should be very close to a QTL to experience hitchhiking, since selective sweep occurs at a very short distance.

Throughout the chapter, modelling was confirmed to be an efficient approach to make useful predictions of the evolution of selected populations although the basic assumptions considered in the models (polygenic additive model, normality of the distribution, base population at the equilibrium, etc.) are not met in reality.

The effective size (*Ne*) of the population is a key parameter for estimation of genetic variability and was estimated using either pedigree information or variance in allele frequencies over time in the experimental chicken lines. Estimated effective size of the population based on the pedigree approach was always lower than estimated effective size based on the temporal variation approach, whatever the loci that were considered: candidate gene, supposedly neutral or selected markers. But, estimation using genotype information from supposedly neutral marker was lower than estimation using genotype information from markers under selection: genetic diversity in regions under selection is weaker than those of the whole genome.

Chapter 5 was not a direct analysis of genetic diversity but an evaluation of risks consequently to reduction of genetic diversity. The study focused on abnormalities in populations that have experimented a strong bottleneck, such as the French Holstein dairy cattle population. Through simulations, we showed that appearance of genetic defects was due to the reduced

and unbalanced use of bulls. We also investigated the consequences of counter-selection against the deleterious alleles identified in short and long-term.

Résumé

L'objectif de cette thèse est double : examiner en détails d'une part, les liens entre les différents critères obtenus à partir de différents types d'information, tels que les marqueurs moléculaires, les pedigrees ou les phénotypes des caractères quantitatifs, et d'autre part, l'évolution conjointe de la variabilité neutre et de la variabilité sélectionnée.

En fonction de la nature des informations, la variabilité génétique et son évolution sont décrites par différents paramètres qui s'appuient sur des modèles sous-jacents plus ou moins complexes et la représentation qu'on se fait de la diversité génétique sera donc plus ou moins réaliste. Ainsi, l'analyse du polymorphisme nous donne directement une image précise de la diversité au niveau du génome. Par exemple, les génotypes et les fréquences alléliques de gènes candidats, ou plus généralement de marqueurs moléculaires, nous donnent directement accès au polymorphisme spécifique à des gènes identifiés en des endroits précis du génome ; quant aux analyses basées sur les pedigrees, elles traduisent le polymorphisme en un locus neutre et anonyme, n'importe où dans le génome. Par contre, l'analyse des phénotypes nécessite des modèles de représentation plus complexes et nous une vue d'ensemble générale de la diversité génétique.

Les principales méthodes pour l'analyse de la diversité génétique à partir d'informations de natures différentes, et celles utilisées au cours de cette thèse, sont présentées dans le premier chapitre.

Les chapitres 2, 3 et 4 constituent une valorisation des différents types d'informations issues de lignées expérimentales de poulets sélectionnées pour des critères de réponse immunitaire. L'ensemble de ces analyses constitue une analyse intégrée de la diversité génétique.

Dans le chapitre 2, nous analysons la diversité génétique et le gain génétique, sur la base des performances et des données généalogiques. Les valeurs observées de consanguinité et de gain génétique sont confrontées à des valeurs prédites à partir de modèles théoriques : les méthodes déterministes ont permis d'obtenir des valeurs prédictives proches des valeurs observées, les différences étant principalement dues à l'écart entre les hypothèses sur le schéma de sélection et la réalité, et les simplifications mathématiques faites dans les modèles de prédiction. Nous avons également testé l'effet du schéma de sélection sur la consanguinité et les autres critères de diversité génétique obtenus à partir des généalogies. Le nombre

efficace d'ancêtres s'est révélé être le paramètre le plus pertinent pour suivre l'évolution de la diversité génétique puisqu'il prend en compte la perte de diversité qui a lieu au cours des générations sous l'effet de la dérive génétique.

L'estimation des paramètres génétiques et leur évolution sont présentées au chapitre 3, ainsi que l'évolution du polymorphisme d'un gène candidat. Afin de tester la neutralité du CMH sur les caractères de nos lignées sélectionnées, nous avons analysé les composant de la variance et avons également comparé l'évolution du polymorphisme dans les lignées avec l'évolution théoriques issues de modèles de prédiction. Au-delà de la mise en évidence de l'effet du CMH dans la réponse immunitaire, cette étude a permis de souligner l'intérêt de combiner différentes approches afin de juger de l'effet d'un gène candidat, et de son évolution, et ce plus particulièrement quand ce gène présente des allèles rares. Nous avons également examiné les variations de la diversité génétique additive au cours du temps, en tentant de surmonter les difficultés pour estimer les paramètres génétiques liées au modèle théorique sous-jacent (modèle polygénique infinitésimal). Pour cela, nous avons testé l'augmentation du nombre de générations disponibles pour l'analyse par REML, et l'utilisation de l'information de quelques générations seulement ; ces approches se sont avérées satisfaisantes pour observer l'évolution de la variance génétique additive et inférer sur l'effet de la sélection sur la réduction de la variance génétique additive.

Dans le chapitre 4, nous avons analysé et comparé l'évolution du polymorphisme de marqueurs moléculaires supposés neutres ou soumis à sélection. Afin de mettre en évidence la signature de sélection laissée par les QTL, nous avons combiné différentes méthodes, simulations et méthodes statistiques. L'étude a montré l'intérêt de comparer l'image de la diversité génétique définie par le polymorphisme de marqueurs situés dans des zones QTL à celle définie par le polymorphisme des marqueurs supposés neutres.

Au travers des différents chapitres, la modélisation est apparue comme une approche efficace pour prédire l'évolution des populations sélectionnées, même si les hypothèses de modélisation faites (telles que le model additif polygénique, la normalité des distributions ou la population des fondateurs sous équilibre de Hardy-Weinberg, etc.) ne correspondent pas exactement au réel modèle biologique.

La taille efficace de la population (Ne) est un paramètre clé dans l'estimation de la variabilité génétique que nous avons estimé dans les lignées expérimentales de poulets, soit à partir des généalogies, soit à partir des variations des fréquences alléliques. Il est apparu que les estimations faites à partir des généalogies étaient toujours inférieures aux estimations faites à partir des fréquences alléliques, quelque soit le type de locus considéré (gène candidat,

markers supposé neutre ou soumis à sélection). L'estimation de Ne à partir des fréquences alléliques des marqueurs neutres a donné une valeur plus faible que celle à partir des fréquences alléliques des marqueurs soumis à sélection. Ceci nous permet de dire que la diversité génétique est plus réduite dans les régions soumises à sélection que dans le reste du génome.

Au chapitre 5, nous n'avons pas analysé la diversité génétique en elle-même mais les conséquences d'une réduction de cette diversité génétique. L'étude a porté sur anomalies génétiques dans les populations ayant subi un fort goulot d'étranglement, telles que la population française bovine laitière Prim'Holstein. Nous avons montré que l'origine des pics de veaux mort-nés venait du très faible nombre de pères et de leur utilisation non équilibrée. Nous avons aussi examiné les conséquences, à court terme et à long terme, d'une contre sélection exercée sur l'allèle délétère responsable d'une anomalie.

Table of contents

INTRODUCTION

During the last decades, agriculture in Europe tended towards intensive and specialised production systems. Up to the seventies, financial availabilities and public policies have supported the development and use of a small number of specialised breeds. Consequently, a limited number of specialized breeds have extended at the national and even worldwide levels since their efficiency has been largely improved (*e.g.* Prim'Holstein in dairy cattle, Large-White in pigs or specialised strains in poultry). Meanwhile, a large proportion of local breeds declined, kept with rather small numbers, some of them are now even endangered or have already disappeared. Then, the question rose about preservation of a breed when this breed that is native to a particular area appears to have lost its function in that area and is consequently in danger of becoming extinct. Therefore, since the 70's conservation programmes were set up in France with the financial and technical support from the State and the European Union. Objectives of the programmes diverged according to the status of the populations: for endangered breeds, objectives were to avoid extinction whereas objectives were to preserve local breeds by developing their values for animal production or environmental purposes.

The evolution and the current status of French cattle breeds may illustrate these trends (Verrier *et al.*, 2001). Figure 1 shows the evolution over time of the repartition of the French cattle stock into different categories of breeds.

During the second part of the XXth century, there was a regular increase of the proportion of cows belonging to the three main dairy breeds (namely, the Prim'Holstein, Montbéliarde and Normandy breeds) or to the three main beef breeds (namely, the Charolaise, Limousine and Blonde d'Aquitaine breeds). Note that the decrease of the part of dairy breeds and the increase of the part of beef breeds observed from the 80s' is a consequence of the European milk quota policy started in 1984. On the contrary, the proportion of cows belonging to other breeds (mainly local breeds) decreased from 35 % in 1943 to only 6 % in 2000. For some local breeds, the breeders succeeded in stopping the decrease of the population size and founded the conditions of a new development of their breed, by developing typical and high quality products under labelling procedures, e.g., PDO cheeses for the Abondance and Tarentaise breeds in the Northern Alps (Verrier *et al.*, 2005 *a*), high prices calves for the Aubrac breed, by using crossbreeding with Charolais bulls in the Massif Central Mountains (Lambert-

Derkimba *et al.*, 2006). Finally, a number of rare breeds benefited of successful conservation programmes (Avon *et al*., 2006).

Arguments in favour of preservation are that we do not know what type of animals will be required in the future and that we should preserve the available genetic diversity between breeds, *i.e.* biodiversity, as an insurance against the unknown future. To decipher which populations to be preserved, it was proposed to analyse genetic distances between populations or the marginal contribution of each population to the diversity of a set of populations (Thaon d'Arnoldi *et al.*, 1998). But use of genetic distances for AnGR management has been criticised (Ruane, 1999) and Weitzman's diversity function (1992) is a method based on pairwise genetic distances between the populations that does not take within-population genetic diversity. Fabuel *et al* (2004) proposed a methodology that allows estimating loss or gain of genetic diversity if removing a population of the set and that takes both within and between population diversity. From a strict genetic point of view, a most important question lies in the information used to analyse the diversity and to take conservation decisions: if the purpose of conservation is insurance for the future, the neutral diversity cannot be the only criterion and the functional diversity is to be included in the analyses.

Arguments for preservation may be also applied to larger selected populations since these breeds show narrow genetic basis, too. For instance, in order to carry on with the French dairy cattle example, it was shown (Mattalia *et al.*, 2006) that half the gene pool of the cows born in 2000-2003 originated from only 8, 8 and 9 major ancestors for the Prim'Holstein, Montbéliarde and Normande breeds, respectively. Other examples, in several livestock species, show that intensively selected breeds with a large population size may be small populations from a genetic point of view (for a synthesis, see Verrier *et al.*, 2005 *b*). Thus, procedures for management of genetic variability are needed in these cases too, with the objective to combine high genetic gains and preservation of the genetic variability.

In fact, characterisation and management of Animal Genetic Resources (AnGR) are essential for sustainable breeding of animal populations as well as answering demands that arise at different level of the production chain: demands from producers but also from the public and consumers themselves. First, genetic variability is necessary for the sustainability and competitiveness of animal productions, by allowing genetic progress concerning production traits that reduce costs and increase quality of the products. This genetic progress should also

concern traits that were not selected yet and deal with animal welfare and consumer health, due to the social demand. Second, local breeds may present characteristics, such as robustness, that could be useful if changes occur in production systems, especially if systems tend to be less intensive. Moreover, local breeds carry both cultural and historical dimensions which are part of the identity of countries.

Genetic diversity can be observed both within and between breeds. But genetic diversity is more obvious in differences between breeds and most of the studies published for both natural and selected populations are focused mainly on the variability among breeds, populations or flocks. For instance, Laval *et al.*, (2000) described genetic variability among European pig breeds (large selected breeds as well as local breeds). The number of such studies is so large that it would not be possible to be exhaustive. However, the FAO has published a review on recent diversity studies in order to evaluate the current status of molecular genetics research in species of domestic animals (FAO, 2004). Yet management of within genetic diversity is a burning issue for selected populations.

Several methods have been proposed to analyse the genetic variability of animal populations that are based on different available kinds of information, such as molecular markers, pedigrees, or phenotypes for quantitative traits.

When the genetic determinism of the observed trait is based on a single locus, variability can be easily determined according to Mendelian rules, but when the genetic determinism is more complex, statistical analysis are needed for estimation of genetic parameters, often based on a model that defines additive genetic effects for all animals individually and accounts for variances covariances among them, the 'animal model' (Bovenhuis *et al.*, 2002).

Knowledge of pedigree allows calculating inbreeding and kinship coefficients and probabilities of gene origin (Boichard *et al.*, 1997), *i.e.* polymorphism of a hypothetical neutral locus, randomly chosen in the genome. This method has been used for various species: sheep as well as for horses or dogs (Huby *et al.*, 2003; Moureaux *et al.*, 1996; Leroy *et al.*, 2006).

Molecular markers are commonly used to describe within genetic diversity using their allelic richness or criteria based on their allele frequencies, such as expected heterozygosity under Hardy-Weinberg equilibrium or effective size of alleles. Molecular information may be also

useful as additive information to pedigree information when parentage assignment is not clearly assessed, as for fish populations (Estoup *et al.*, 1998).

Though, numerous methods aim at the optimisation and the management of within genetic diversity in selected populations.

In small sized selected populations, management of genetic diversity can either be done using pedigree or molecular data. On one hand, applying simple demographical rules were found to be very efficient to restrict the rate of inbreeding such as using as many males as possible in order to balance progeny sizes, quickly replacing old sires by new ones and splitting populations into some reproduction groups: this efficiency was assessed by theoretical arguments (e.g., Hill, 1972), simulation studies (Rochambeau and Chevalet, 1985) or *a posteriori* analyses of real populations, such as the Solognot sheep breed which is under conservation (Huby *et al.*, 2003).

In the case of overlapping generations, Nomura (2005) proposed and tested through simulations methods for minimizing loss of genetic diversity by minimization of the average coancestry either in the next generation or on long-term, both methods relying on gene flow.

On the other hand, use of molecular information is another option in conservation programmes, by monitoring evolution of genetic variability based on criteria calculated using allele frequencies or by choosing reproducers and planned matings according to individual information from certain markers of interest. For instance, reproducers may be chosen one by one or by groups considering their heterozygosity rate or them carrying a rare allele.

However, combining both sources of information, pedigree records and molecular markers, should be favoured in conservation programmes unless molecular information is available for a large number of individuals (Fernandez *et al,* 2005): then, effective size of the population clearly increases.

For larger selected populations, very promising methods are based on a joint optimisation of the average genetic value and the average relationship or inbreeding in each new-born generation: Meuwissen (1997) maximise the genetic level of selected animals while constraining their average coancestry to a predefined value whereas Colleau *et al.* (2004) minimize the average coancestry while constraining average breeding value set to a desired value.

Optimisation method developed by Colleau *et al.* (2004) was tested in selected French pig and dairy cattle populations (Colleau and Moureaux, 2006; Colleau and Tribout, 2006): retrospective optimisations for the planned matings in the French Landrace population and in

the Prim'Holstein, Montbéliarde and Normande breeds were carried out. For both examples, the major constraint was that the average overall estimated breeding value should be the same as observed, in order not to damage short-term genetic gain, and optimisations of reproducers' contributions have led to decrease coancestry and inbreeding coefficients by 20 to 30% in comparison to the actual matings.
Optimisation are based on relationships obtained from pedigree, assuming that genetic models of independent loci. However, Fernandez *et al.* (2000) have shown that linkage between the markers could reduce efficiency of these optimisation methods and the use of pedigree information was found to be useful unless the selected trait was controlled by genes spread in large genome regions.

Yet do we have a clear picture of the within genetic diversity in domestic animal populations? Is this picture the same according to the kinds of information considered or according to the tools and methods applied, and thus the underlying model? Does the picture we get at the individual level may be generalised to the whole population? And the other way around? Has selection (directional or artificial) equally affected the genome?
Moreover, effective size of the population seems to be a key parameter in analysing and comparing the within- and between population genetic diversity and this parameter may be estimated through various methods, using different kinds of information as well. Then, focusing on this particular genetic parameter of diversity should be interesting.

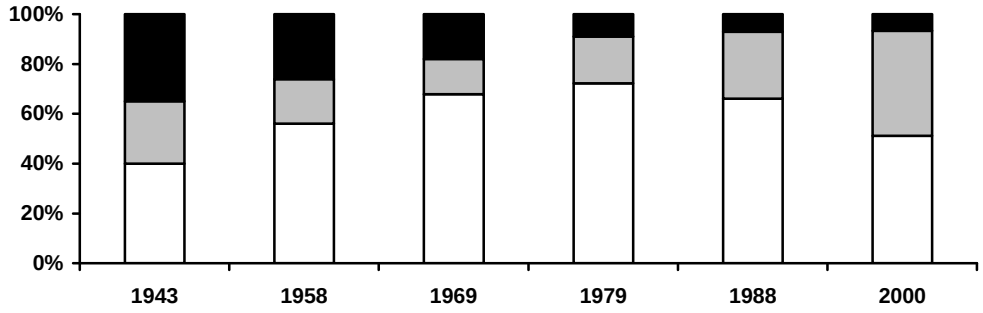
Therefore, in order to answer those questions, the scientific purpose of the thesis is twofold: (*i*) to go further in the examination of the joint evolution of the neutral variability and the selected variability, (*ii*) to investigate in details the links between criteria originating from different kinds of information.

This thesis is mainly concerned with results from chicken experimental lines but also analyses dataset from dairy cattle population; the objective is not to restrict results to those two species but to be able to draw guidelines for general use in analysing genetic diversity.

The thesis was divided into five chapters: the first chapter introduces the methodology whereas each of the four following chapters focuses on distinct aspects of genetic diversity. Chapter 2 and 3 deal with a valorisation of different kinds of information from experimental chicken lines for the integrate analysis of genetic variability: chapter 2 deals with an analysis

based on performances and pedigree data whereas chapter 3 deals with estimation of genetic parameters and polymorphism evolution of a candidate gene. Chapter 4 is an analysis and comparison of the polymorphism evolution of molecular markers, some markers being supposed neutral, the others supposed to be under selection, in chicken experimental lines. Finally, chapter 5 is an evaluation of risks consequently to reduction of genetic diversity due to bottlenecks in large selected populations.

Figure 1: Evolution in size of the French cattle breeds: three main dairy cattle breeds (white), three main beef cattle breeds (grey) and other breeds (black).



[*Source*: Ministry of agriculture, Agricultural census]

CHAPTER 1: Methods applied for the analysis of genetic diversity

Tools that are available to monitor genetic diversity are based on different kinds of information, such as genetic markers, especially the molecular markers, pedigrees, or phenotypes for quantitative traits. According to the kind of information, genetic diversity and its evolution will be described through various parameters dealing with more or less complex underlying model, thus, we will have more or less realistic representation of the genetic diversity. On one hand, analysis of polymorphism gives a direct overview of the diversity: genotypes and allele frequencies of candidate gene or molecular markers will give access to specific polymorphism of known areas in the genome, whereas pedigrees will give access to anonymous polymorphism of unknown areas in the genome. On the other hand, analysis of phenotypes gives a general overview of genetic diversity, assuming the model of representation to be more complex.

Genetic diversity can be observed both within and between breeds but we will mainly focus here on methods that apply to analysis of the within genetic diversity. In this chapter, the main methods available for the analysis of genetic diversity using different kinds of information are introduced, and the methods that will be used furthermore in the thesis are presented.

1. Polymorphism at candidate genes or molecular marker

1.1. Current indicators and heterozygosity

The first indicator to describe a given locus is its total number of alleles (allele richness). This parameter may often be underestimated if the sampling size is small but polymorphism is large (for instance microsatellite loci); then, rare alleles may never be sampled. When dealing with several populations and different sample sizes, some methods are available to let homogeneous the estimations of allelic richness (see Foulley and Ollivier, 2006). However, this indicator only gives a partial and static overview on diversity, since it does not take allele frequencies into account.

Several indicators based on allele frequencies have been proposed. Heterozygosity is of major interest in studying genetic variation in populations. Several measures of heterozygosity exist,

such as expected heterozygosity (H_{exp}, or gene diversity, *D*, as Weir prefers to call it). For a single locus it is calculated as: $H_{\exp} = 1 - \sum_{i=1}^{k} p_i^2$, where p_i is the frequency of the i^{th} of *k* alleles. H_{exp} represents the probability of having two different alleles at a given locus when drawing at random two genes in a population. Similarly, the effective number of alleles (*Ae*) is defined as the reciprocal of the probability that two genes drawn at random are the same allele. For a single locus, it is computed as: $Ae = 1 \Big/ \left(\sum_{i=1}^{k} p_i^2 \right)$, wit the same definitions as above, and then, $Ae = 1/(1 - H_{\exp})$.

When considering several loci, the overall heterozygosity is equal to the average value across loci and then may be computed as follows: $H_{\exp} = 1 - \frac{1}{m} \sum_{l=1}^{m} \sum_{i=1}^{k} p_i^2$, where the first summation is for the l^{th} of *m* loci.

For small samples, Nei (1978) proposed a formula that takes into account the sampling errors:

$$H_{\exp} = \left[\sum_{l=1}^{m} \frac{2n_l (1 - \sum_{i=1}^{k} p_i^2)}{2n_l - 1} \right] \Big/ m$$

where n_l is the number of genes that has been analysed for the l^{th} locus (twice the number of individuals).

The level of heterozygosity to what we expect under Hardy-Weinberg equilibrium is compared to observed heterozygosity. If the observed heterozygosity is lower than expected, the excess of homozygous invoke forces such as inbreeding or subpopulation structure, known as Wahlund effect (Wahlund, 1928). On the contrary, if observed heterozygosity is higher than expected, the excess of heterozygotes may suggest an isolate-breaking effect (the mixing of two previously isolated populations) or assortative mating.

→ *Population structure (between populations variability):*

F-statistics are a set of tools proposed by Wright (1921, 1969) to partition heterozygote deficiency into within and among population components. *Fis* measures the heterozygote deficit within populations, *Fst* among populations and *Fit* the global deficit of heterozygotes.

Estimation of F-statistics have been developed by Nei (1975), *Gis* and *Gst* being estimators of *Fis* and *Fst*, respectively, and by Weir and Cockerham (1984), *sigb* and *siga* being estimator of *Fis* and *Fst*. Estimators differ on some points: for instance, Nei's statistics weight all samples equally, whatever the sample size whereas Weir and Cockerham's statistics weight allele frequencies according to sample size. They also treat differently monomorphic loci: Weir and Cokerman consider that estimators can not be defined whereas *Gis* and *Gst* are set to zero.

Rst is an estimator of gene differentiation that accounts for variance in allele size and is defined for markers undergoing a stepwise mutation model (Slatkin, 1995). However, this estimator should be used unless mutation can be neglected compared to other forces or mutation follows a stepwise mutation strictly.

→ *Within population variability and its evolution:*

In populations of finite size, the effect of genetic drift results in a change in allele frequencies, which results in a decline in heterozygosity at a rate of 1/2*Ne* ,where *Ne* is the effective size of the population. Let H_t the average heterozygosity for a given biallelic locus (allele frequencies *p* and 1-*p*) at the t^{th} generation, then $H_t = 2p(1-p)(1-\frac{1}{2Ne})^t$. If the initial population (generation 0) is produced by random mating, the expected heterozygosity at that time is supposed to be 2*p*(1-*p*), then $H_t = H_0(1-\frac{1}{2Ne})^t$. When the effective size is large enough, we may approximate as follows: $H_t = H_0 \cdot e^{-\frac{t}{2Ne}}$

This stochastic change in allele frequencies (called genetic drift) resulting simply from the finite size of a population implies that the smaller the population size (*Ne*), the faster the decline in heterozygosity.

1.2. Temporal variation of allelic frequencies

The above parameters may be calculated from samples at different generations and be compared in order to analyse evolution of the within genetic variability. Another approach consists in directly estimating the temporal variation of allele frequencies and to compare the

observed values to expected values drawn under various hypotheses in order to distinguish the different forces applied on the evolution of polymorphism.

Effective population size may be indirectly estimated by measuring temporal changes in allele frequencies (*f*), also called variance effective size (Ne_V). Several methodologies have been developed either based on a temporal approach or on a likelihood approach.

→ *Temporal-based approach:*

Authors studying the temporal method for estimation of *Ne* (Nei and Tajima, 1981; Pollak, 1983) assumed a diploid, random mating population of size *N*, from which samples were drawn at generation 0 and at generation *t* (sample of S_0 and S_t individuals, respectively). Generations were supposed to be discrete, and selection, migration and mutation forces were presumed to be negligible. Nei and Tajima (1981) proposed an estimator of *f* for a single locus, as follows:

$$\hat{f}_C = \frac{1}{k}\sum_{i=1}^{k}\frac{[x_i(t)-x_i(0)]^2}{\frac{x_i(t)+x_i(0)}{2}-x_i(t)\cdot x_i(0)}$$

where *k* is the number of segregating alleles, $x_i(0)$ and $x_i(t)$ are the allelic frequencies at generation 0 and at generation *t*, respectively. Pollack (1983) proposed a variant:

$$\hat{f}_K = \frac{1}{k-1}\sum_{i=1}^{k}\frac{[x_i(t)-x_i(0)]^2}{\frac{x_i(t)+x_i(0)}{2}}$$

For multiple loci, weighted means of the single locus $\hat{f}$ are computed.

Smaller estimates of *f* lead to larger estimates of *Ne*, and $\hat{f}_K$ is larger than $\hat{f}_C$ because it has smaller denominator. However, on simulation basis, Waples (1989) showed that the choice of which estimator of *f* to use had small effect on estimation of *Ne*, both mean $\hat{f}_C$ and mean $\hat{f}_K$ being accurate, although both tended to overestimate *Ne* slightly. Because of its larger value, $\hat{f}_K$ led to lower estimates of Ne but $\hat{f}_C$ was slightly more accurate for loci with unbalanced allelic frequencies. Therefore, we will consider estimator given by Nei and Tajima (1981) along this thesis.

Two different sampling plans (individuals for genetic analysis) were identified (Nei and Tajima, 1981): in plan I, individuals are taken after reproduction or are replaced before reproduction occurs, whereas in plan II, individuals are taken before reproduction and not

replaced. According to sampling plan, Waples (1989) gave a different estimator of effective size population. In plan I, *Ne* may be estimated as:

$$\hat{N}e_V = \frac{t}{2\left[\hat{f} - \frac{1}{2S_0} - \frac{1}{2S(t)} + \frac{1}{N}\right]}$$

whereas in plan II, estimator is:

$$\hat{N}e_V = \frac{t}{2\left[\hat{f} - \frac{1}{2S_0} - \frac{1}{2S(t)}\right]}$$

where $\hat{f}$ is stands for estimator of variation in allele frequency change, either $\hat{f}_C$ or $\hat{f}_K$.

As pointed out by Nei and Tajima (1981), precision for the temporal method increases with the ration *S*/*Ne*, which means that populations with small Ne are most effectively analysed.

Sampling plan II is the sampling plan to be considered along this thesis.

Distribution of *Ne* is far from normal since the distribution $\hat{f}$ (on which *Ne* is based) is skewed. This is why, Nei and Tajima (1981) considered that variance of $\hat{f}$ and Ne were not suitable for calculating confidence intervals (CIs) and suggested using the 2.5 and 97.5% cumulative probabilities of the χ^2 distribution:

$$95\%\ \text{CI for}\ \hat{f} = \left[\frac{\chi^2_{0.975[n]}\hat{f}}{n}; \frac{\chi^2_{0.025[n]}\hat{f}}{n}\right]$$

However, this method is only correct in the case of large number of alleles. Considering that $\hat{f}$ is calculated on a finite number of alleles, as it will be the case thereafter in the thesis when using microsatellites data, and that *f* is actually a variance of allelic frequencies, formula for the CI of a variance may be applied (Waples, 1989):

$$95\%\ \text{CI for}\ \hat{f} = \left[\frac{n \cdot \hat{f}}{\chi^2_{0.975[n]}}; \frac{n \cdot \hat{f}}{\chi^2_{0.025[n]}}\right]$$

Bounds for the estimate of *f* can be used to calculate bounds for the estimate of *Ne*.

Calculated *f* is compared with the distribution of *f* obtained from a series of simulations of a population undergoing pure drift and that presents the same initial allelic frequencies and the same inbreeding effective size (Goldringer and Bataillon, 2004). This test allow us to identify loci with extreme *f* values compared to the rest of the genome: such loci are likely to be in areas undergoing selection, driving *f* to values greater then expected values under drift alone.

The temporal method is affected by bias due to the small size of samples or due to highly skewed allele frequencies. In our different studies, size of the samples will not be a problem; however, microsatellite allele frequencies appear to be unbalanced (as well as SNPs), inducing a downward biased estimation of *f* and, as a consequence, to an upward bias of *Ne*. In order to overcome the different sources of bias affecting the temporal method, Jorde and Ryman (2007) proposed an tested through simulations an alternative estimator of *f*, noted F_s, that weights alleles according to their heterozygosity: a lower weight is given to alleles with low frequencies in the samples, so that they less contribute to the mean estimate, in the same way Reynolds *et al.* (1993) and Weir and Cockerham (1984) recommended to weight each allele when estimating genetic distances. This new estimator was not used in our studies, since the analyses were previously undergone, but would have certainly given a more accurate estimator of the variance effective size.

→ *Likelihood-based approach:*

With a likelihood-based approach, the role of the observed values and the distribution parameters are reversed. Thus, we estimate here the value of the effective size population that is likely to produce the observed allele frequencies.

Berthier *et al.* (2002) proposed a likelihood-based estimator that uses two temporally spaced genetic samples of individuals from a population, based on a coalescent approach. Estimation of the effective population size was similar to the estimation using the moment-based method (with larger confidence intervals in the case of the likelihood-based method), when a small population size was considered (Berthier *et al.*, 2002).

Williamson and Slatkin (1999) first proposed a maximum-likelihood method where the data are collected at different generations. The first sample was collected at generation 0 and the other samples were drawn at successive generations, evenly or irregularly spaced in time. In simulation studies, Williamson and Slatkin (1999) showed that maximum-likelihood estimator outperformed the moment-based estimators: the maximum-likelihood estimator had a lower variance and a smaller bias, although both maximum-likelihood-based method and the moment-based method overestimated the effective population size. However, this method was restricted to data on biallelic loci for computational reasons in estimating the likelihood of *Ne*. This drawback was fulfilled by Anderson *et al.* (2000) who presented a Monte Carlo approach to compute the likelihood with data on multi-allelic loci.

Excluding computational considerations, we have chosen not to apply the method proposed by Anderson *et al.* (2000) because samples were drawn in two extreme generations (Generation 0 and generation 11) for microsatellite markers (see Chapter 4). Although genotypes at MHC locus were known at each generation for all individuals, we have chosen not to apply this method in the case of MHC locus (see Chapter 3) in order to be able to compare the polymorphism evolution of this candidate gene and those of the microsatellites on the same basis.

1.3. Genotypic linkage disequilibrium

Presence of genotypic disequilibrium in data set may be tested using the chi-square test for the difference between the observed frequencies and the expected frequencies under Hardy-Weinberg equilibrium. But when genotypes are scored, it is often not possible to distinguish between the two double heterozygotes (for instance, AB/ab and Ab/aB), so that the gametic frequencies cannot be inferred. Then, a contingency table based on the observed genotypic data is generated for each pair of loci. Testing for genotypic disequilibrium means testing the significance of association between genotypes at pair of loci in a sample. The statistics used to test the tables is the log-likelihood ratio G-statistic (the only part of the statistic that changes with randomizing tables): $\sum_{i<=j}\sum_{k<=l} x_{ijkl}\ln(x_{ijkl})$

where x_{ijkl} represents the number of individuals in the sample with genotype *ij* at the first locus and genotype *kl* at the second locus. The test determines whether the resolutions of haplotypes are significantly non-random, which is equivalent to testing whether there is statistically significant linkage disequilibrium between loci.

2. Parameters based on pedigree data

2.1. Inbreeding

One individual is inbred if its parents are related; two individuals are related if they have a common ancestor. Then, the inbreeding coefficient of an individual is the probability that, at a given neutral locus, this individual owns two alleles that are identical by descent.

The simplest definition of inbreeding would be that inbreeding is the mated of related individuals. Because all individuals in a population are related at some degree, we should say that inbreeding is the mating of individuals more closely related than average of the population. The main effect of inbreeding is the increase in the number of homozygous loci in inbred animals and an increase in the frequency of homozygotes genotypes in the population; therefore, inbreeding leads to uniformity through erosion of polymorphism, *i.e.* loss of genetic diversity.

As a consequence, inbreeding increases the chance of deleterious alleles to become homozygous and to express themselves: expression of deleterious alleles with major effects, particularly lethal genes, is the very visible consequence of inbreeding. Expression of unfavourable recessive alleles influencing polygenic traits is not so obvious but if you sum the small effects of those genes altogether, performances may be influenced and even decreased: this phenomenon is known as inbreeding depression.

In populations of finite size, there is sampling in the gene pool from one generation to the next one: genetic drift induces reduction of polymorphism and increase of inbreeding. Then, monitoring increase of inbreeding is a good tool for managing genetic diversity in those populations.

In the present thesis, individual inbreeding coefficients were calculated using the algorithm of van Raden (1992), derived from the tabular method. Its principle consists in building the relationship matrix of each individual and its ancestors; the additive relationship matrix is a symmetrical matrix relating all individuals in the pedigree. Inbreeding coefficient of an individual being equal to the coefficient kinship of its two parents, inbreeding coefficients were deduced from the diagonal elements of the relationship matrix and kinship coefficients between two individuals were deduced from the elements, multiplied by ½, of the relationship matrix out of the diagonal. Let assume F_i the inbreeding coefficient of an individual i, $\Phi_{p,m}$ the kinship coefficient of its parents p and m, and the additive relationship matrix ($\boldsymbol{A}$), then

$F_i = \boldsymbol{A}_{i,i} - 1$ and $\Phi_{p,m} = ½\, \boldsymbol{A}_{p,m}$.

The rate of inbreeding (ΔF) from one generation t to next generation t+1 is defined by: $\Delta F = (F_t - F_{t-1})/(1 - F_{t-1})$ (see, for example, Falconer and Mackay, 1996).

Rearrangement of this equation over generations leads to: $1 - F_{t_0+t} = (1 - \Delta F)^t \cdot (1 - F_{t_0})$

Hence, assuming a constant rate of inbreeding across generations, ΔF is estimated according to the equation: $\Delta F = 1 - \sqrt[t]{(1 - F_t)/(1 - F_0)}$

Inbreeding effective size (Ne_I) was estimated from the rate of inbreeding, according to the classical formula: $\hat{N}e_I = \frac{1}{2 \cdot \Delta F}$

2.2. *Probabilities of gene origin* (James, 1972; Boichard *et al.*, 1997)

When a gene is randomly sampled at any autosomal locus from any given animal, probability of coming from its sire is 0.5 and probability of coming from its dam is 0.5. This rule is applied on the complete pedigree and leads to the probability that the gene comes from one of its founders, i.e. probability of gene origin, o also called long-term genetic contribution

- *Founders and effective number of founders*

A founder is defined as an ancestor with unknown parents. Its expected contribution (r) is the probability that a gene randomly sample in the population under study (called "reference population**"**) comes from this founder. By definition, the founders contribute to this population without redundancy and expected contributions over all founders sum to one.

If the number of founders is f, then: $\sum_{k=1}^{f} r_k = 1$

The effective number of founders f_e is the reciprocal of the probability that two genes drawn at random in the reference population come from the same founder. This parameter is estimated as: $f_e = 1/\sum_{k=1}^{f} r_k^2$

and can be interpreted as the number of founders that would be expected if they would equally contribute to the reference population and lead to the same genetic variability (Boichard *et al*., 1997).

- *Major ancestors and effective number of ancestor*s

The effective number of ancestors is a direct analogy of the effective number of founders: all generations are taken into considerations instead of only the generation of the founders.

As ancestors may not be founders and may be related, their expected contributions may be redundant and their sum may exceed one.

A major ancestor is an ancestor with a high-expected contribution. Therefore, when a major ancestor is identified:

1. Its pedigree information is deleted and it becomes a “pseudo-founder”: family links with its dam and with its sire are deleted;
2. Its contribution is then called “marginal contribution”, i.e. “the contribution not explained by the other ancestors”;
3. Contributions of other individuals are updated:

If one of the parents (sire or dam) contributes to the reference population only through this major ancestor, then the contribution of this parent equals zero; else, the contribution of this parent is updated (decreased) and the contributions of individuals related to this parent are consequently updated, too.

Major ancestors are chosen one by one in an iterative procedure.

3. Components of variance for a quantitative trait

Animal breeding programs are based on the principle that the phenotype of an individual provides some insight into its underlying genetic value; therefore, representation models are built to be able to study complex character, assuming that traits may be analysed without reference to specific genes and variations may be explained by the multiple-factor hypothesis.

The performance (P) of an animal, also called phenotypic value for a given trait, is assumed to be the sum of the mean value (μ), an additive genetic value (A) and an independent residual value (E). This residual value includes both non additive genetic effects and environmental effects. However, in most of the models used, the residual value includes environmental effects only. Then, the phenotypic variance (σ^2_P) is assumed to be the sum of the genetic variance (σ^2_A) and the environmental variance (σ^2_E): $\sigma^2_P = \sigma^2_A + \sigma^2_E$. The underlying genetic model is assumed to be an infinitesimal additive polygenic model: the trait is determined by a very large number of loci with small effects, effects of alleles at each locus being additive.

In random mating populations of finite size, with an additive genetic model, the additive genetic variance (σ^2_A) is assumed to decline over time, due to the genetic drift. With directional selection, additive genetic variance is assumed to decline due to both the drift and the generation of gametic disequilibrium (Bulmer, 1971).

Knowledge of genetic parameters for animal breeding, such as additive genetic variance, heritability (in its narrow-sense, *i.e.* ratio σ^2_A / σ^2_P) and genetic correlations are needed to predict response to selection and to predict breeding values of candidates for selection.

The simplest design for estimation of genetic parameters is the parent-offspring regression: regression of offspring phenotypes against phenotype of one of its parent, the slope of the regression being the direct estimate of twice the heritability.

The use of large populations, with well-balanced designs and within generation selection, allows the estimation of genetic parameters using simple methods, such ANOVA: estimate of σ^2_A may be obtained from the covariance between records of relatives. For instance, in a parent-offspring design, σ^2_A = 2*Cov(parent, offspring) or in a half-sib design, σ^2_A = 4*Cov(half-sib records). The simple full-sib design suffers that it can not give a direct estimation of σ^2_A ; then, a nested full-sib design is commonly used, where each male (sire) is mated with unrelated females (dams), generating a series of nested families nested within half-sibs. The advantage of ANOVA is that estimators are unbiased regardless data are normally distributed or not, but sample sizes have to be well balanced and specific design are not always met in field observations. Henderson (1953) proposed modifications to the ANOVA sum of squares in order to take unbalanced data into account. However, with these quadratic methods, probability of getting estimates outside the parameter space (negative variance for instance) is not null (Searle, 1989).

Unlike ANOVA, Maximum Likelihood (ML) and Restricted Maximum-Likelihood (REML) do not require any specificity about the design or balance of data and are powerful approach to estimating variance components in complex but known pedigrees, based on an animal model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

where **y** is the phenotypic value for a trait, $\boldsymbol{\beta}$ and **a** are the vector of fixed and additive genetic effects respectively, **X** and **Z** are the incidence matrices and **e** is the vector of residual deviations. One drawback of ML approach is that all fixed effects are assumed to be known without error: this is scarcely true in reality and as a consequence, ML estimators give bias estimation of variance components. The advantage of REML is that the method estimates the parameters of the distribution that maximize the proportion of the likelihood of the observed data that does not depend on the fixed effect (such as the effect of sex or year).

REML provides estimations of σ^2_A, σ^2_E and covariances in the base population. If the base population consists of unrelated, unselected and non inbred individuals, then REML gives unbiased estimates of the additive genetic variance in the base population (Gianola and Fernando, 1986).

Yielding unbiased estimates of genetic variance appears to be difficult, due to statistical methodologies undertaken and the assumptions of the underlying model. Actually, van der Werf and de Boer (1990) have used Monte Carlo simulations and shown bias in estimates of genetic variance from data with base animal being selected. Graser *et al.* (1987) suggested to use a model where base animals would be treated as fixed effect but this approach also seemed to be biased (van der Werf, 1992). Moreover, estimation of variance components in quantitative genetic models often involves records on selected animals and condition of including all data on which selection was based cannot always be fulfilled, inducing bias in estimation of genetic parameters. However, Sorensen and Kennedy (1984) have shown on simulation basis that estimates of genetic variance would not be affected by selection over generations if all data and all genetic relationships since the beginning of selection were included in the analysis.

It is with some more difficulties for one yielding estimates of genetic variance over time since REML provides estimation in the base population. But proposals using REML for estimation of additive genetic variance over time are made, such as the examination of Meyer and Hill (1991) of a long-term experiment in mice, using partial data and relationship information: this study has assessed changes in additive genetic variance and covariance during the course of the experiment. This methodology will be tested thereafter in this thesis.

Variance components being estimated in the base population, the BLUP (Best Linear Unbiased Predictor) and BLUE (Best Linear Unbiased Estimator) methods may be used to predict breeding value of each individual and estimate fixed effects: they are linear functions of the observed phenotypes that minimize the sampling variance and are unbiased since $E[BLUE(\beta)] = \beta$ and $E[BLUE(u)] = u$.

CHAPTER 2: Analysis of genetic gain and genetic diversity based on performances and pedigree data

This chapter deals with an analysis of genetic diversity from two independent selection experiments that have been realised in poultry. One experiment was handled by the Animal Breeding and Genetics group of Wageningen University and Research Centre (WUR), the Netherlands, and lines were undergoing true truncation selection. The other experiment was handled by the Animal Genetics and Diversity joint research unit (INRA / AgroParisTech), in its experimental unit "Unité Expérimentale de Génétique Factorielle Avicole", in Nouzilly, France. Lines were undergoing truncation selection with a balanced representation of half-sib families.

Obtaining genetic gains is the first goal of selection in domestic animal populations. However, these gains should not be at the expense of the genetic variability. Then, it is important to predict precisely consequences of selection schemes on diversity parameters. Therefore, the first goal of this study was to validate existing theory by comparing the observed evolution of inbreeding and genetic gain to values predicted by some theoretical models. Second, the study aimed to compare the evolution of some parameters based on pedigree analysis in the different lines according to their management rules.

I. A validation of predictions of genetic gain and inbreeding

(see article)

Genet. Sel. Evol. 37 (2005) 273–289

DOI: 10.1051/gse:2005002

Original article

A comparison of two methods for prediction of response and rates of inbreeding in selected populations with the results obtained in two selection experiments

Valérie LOYWYCK[a*], Piter BIJMA[b], Marie-Hélène PINARD-van der LAAN[a], Johan van ARENDONK[b], Etienne VERRIER[c]

[a] UMR Génétique et diversité animales, Institut national de la recherche agronomique / Institut national agronomique Paris-Grignon, 78352 Jouy-en-Josas Cedex, France
[b] Animal breeding and genetics group, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands
[c] UMR Institut national de la recherche agronomique / Institut national agronomique Paris-Grignon, 16 rue Claude Bernard, 75005 Paris, France

(Received 2 April 2004; accepted 10 November 2004)

Abstract – Selection programmes are mainly concerned with increasing genetic gain. However, short-term progress should not be obtained at the expense of the within-population genetic variability. Different prediction models for the evolution within a small population of the genetic mean of a selected trait, its genetic variance and its inbreeding have been developed but have mainly been validated through Monte Carlo simulation studies. The purpose of this study was to compare theoretical predictions to experimental results. Two deterministic methods were considered, both grounded on a polygenic additive model. Differences between theoretical predictions and experimental results arise from differences between the true and the assumed genetic model, and from mathematical simplifications applied in the prediction methods. Two sets of experimental lines of chickens were used in this study: the Dutch lines undergoing true truncation mass selection, the other lines (French) undergoing mass selection with a restriction on the representation of the different families. This study confirmed, on an experimental basis, that modelling is an efficient approach to make useful predictions of the evolution of selected populations although the basic assumptions considered in the models (polygenic additive model, normality of the distribution, base population at the equilibrium, etc.) are not met in reality. The two deterministic methods compared yielded results that were close to those observed in real data, especially when the selection scheme followed the rules of strict mass selection: for instance, both predictions overestimated the genetic gain in the French experiment, whereas both predictions were close to the observed values in the Dutch experiment.

selection experiments / poultry / inbreeding / genetic response / prediction methods

* Corresponding author: loywyck@inapg.fr

1. INTRODUCTION

Selection programmes are mainly concerned with increasing genetic gain. Short-term progress should not, however, be obtained at the expense of the within-population genetic variability. Conservation of genetic variability should allow breeders to optimise selection programmes in the middle and long-term but also to face biological or economical problems. Selection and drift are two major factors affecting genetic variability. Several mathematical algorithms have been proposed to predict the evolution of inbreeding and/or genetic variability within selected populations of small size [2, 9, 23, 25, 27, 29, 32]. Extensive Monte Carlo simulation work has been done to investigate this evolution in a large range of situations and to compare different methods intended to jointly optimise both genetic gains and restrict the rate of inbreeding, and preserve the variability [11, 17, 28]. The analysis of real populations has mainly concerned the evolution of inbreeding and probabilities of gene origin based on pedigree information [13, 18, 30]. Markers have also been used to assess the evolution of genetic polymorphism [15].

Selection experiments provide the opportunity to check the validity of theoretical models to predict genetic response on a more or less long term [5, 10, 16]. Experimental lines are powerful tools because of the complete knowledge of the pedigree and the existence of control lines and since the environment is controlled. Moreover, laboratory species and some livestock species (*e.g.* poultry) present the twofold main advantage of short and discrete generations.

Theoretical models to predict the evolution within a small population of the genetic mean of a selected trait, its genetic variance and inbreeding have mainly been tested through Monte Carlo simulations. The purpose of this study was to compare theoretical predictions to experimental data. Two deterministic methods [27, 29] will be considered, both assuming the additive infinitesimal model. Therefore, differences between theoretical predictions and experimental results can arise from differences between the true and the assumed genetic model, and from mathematical simplifications applied in the prediction methods. Two sets of experimental lines of chickens were used in this study: one undergoing true truncation mass selection and the other one undergoing mass selection with a restriction on the representation of the different families. Comparisons will be done by analysing response to selection and inbreeding.

2. MATERIALS AND METHODS

2.1. Experimental design and genetic parameters of the selected traits

Two selection experiments on immune response in chickens have been conducted, one in the Netherlands [6] and another one in France [22]. The present study was based on the selected and control lines of these two experiments.

For the French experiment, 9 males and 44 females originating from a White Leghorn line and a commercial line were crossed in 1992. The F1 offspring were randomly crossed to produce the generation 0 (n = 523), from which four lines were derived. Three of these lines were selected for high values, according to three different criteria of immune response: antibody response 3 weeks after vaccination against the Newcastle disease virus (line 1, trait ND3), phagocytic activity at 12 weeks of age (line 2, trait PHA), and cell-mediated immune response at 9 weeks of age (line 3, trait CC). There were two steps in the selection. First, the best animals were selected on the basis of their individual performance. Second, the sizes of the different half-sib families were approximately balanced except that if the performance of a given family was considered as too low, no individuals were selected from this family. As a consequence, this selection was intermediate between strict within-family and mass selection. The fourth line was a control line (line C1), with the parents being chosen at random. In each line and in each generation, 15 males and 30 females were chosen as parents of the next generation out of about 100 candidates in each sex. Mating was at random, except that full and half-sib matings were avoided. This selection design has produced 8 discrete generations (generation 1 to 8) which represent a total number of 6750 measured individuals.

For the Dutch experiment, 24 males and 96 females originating from two ISA Brown chicken lines were randomly crossed in 1981, to produce generation 0 (n = 614), after which selection started. Three lines were derived from generation 0: two selected lines, for high values (line H) and low values (line L), and one control line (line C2). The selection criterion was the individual total antibody (Ab) titre 5 days after injection of sheep red blood cells (SRBC). In lines H and L, a strict mass selection was applied: each generation, 25 males and 50 females were selected out of about 150 candidates in each sex. In line C2, 40 males and 70 females were randomly chosen out of about 125 candidates in each sex. In each line, mating was at random except that full and half-sib matings were avoided. This selection design has produced 18 discrete generations (generation 1 to 18) which represent a total number of 17 194 measured individuals. For each experiment, the heritability

and the genetic variance in generation 0 of each of the selected traits have been previously estimated. The heritability in the base population, taking into account seven generations in the French experiment (generation 0 to 6), was estimated at 0.35 (±0.02) for ND3, 0.13 (±0.02) for PHA and 0.15 (±0.02) for CC [22]. Pinard also showed no significant correlated response and estimated genetic correlations between the three immune traits that did not differ significantly from zero [22]. Therefore, we assume these three traits to be independent. The heritability of the Ab titre, taking into account nine generations in the Dutch experiment (generation 0 to 8), was estimated at 0.29 (±0.05) in line H, 0.36 (±0.04) in line L and 0.22 (±0.04) in line C2 [21].

2.2. Data analysis

At each generation, the cumulated response to selection in a given line was computed as the difference between the phenotypic mean in this line and the phenotypic mean in the corresponding control line, any phenotypic mean being computed over all animals born at the considered generation and measured. The global genetic gain per generation was computed as the slope of the linear regression of the observed cumulated response to selection over generations. The realised heritability was computed as the slope of the linear regression of the cumulated response to selection against the cumulated selection differential.

The individual coefficients of inbreeding were computed from the pedigree relationships using the algorithm of Van Raden [26]. In each line, from generation $t-1$ to generation t, the rate of inbreeding (ΔF_t) was computed following the classic formula (see, for instance, Falconer and Mackay [12]):

$$\Delta F_t = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$$

where F_t is the average coefficient of inbreeding in the considered line, in generation t. Assuming a constant rate of inbreeding across generations, the global rate of inbreeding (ΔF), over t generations, was computed as:

$$\Delta F = 1 - \sqrt[t]{(1 - F_t)/(1 - F_0)}. \quad (1)$$

2.3. Methods of prediction

The observed results on both inbreeding and genetic response in the selected lines were compared to the predictions given by two deterministic methods,

one proposed by Verrier *et al.* [27] and hereafter referred to as the "VCF" method and the other one proposed by Woolliams and Bijma [29] and hereafter referred to as the "WB" method. Both methods assume an additive infinitesimal model with a normal distribution of the genetic and phenotypic values for the trait, random mating and true truncation selection. Prediction inputs are heritability and phenotypic variance of the trait in the base population (assumed here to be generation 0), and numbers of selected and candidate animals in each sex. In the present study, the genetic parameters previously estimated (see Sect. 2.1) were used. From generation t to generation $t + 1$, the response to mass selection (R_t) was predicted according to the classic formula:

$$E(R_t) = i \cdot h_t^2 \cdot \sigma_{P_t} = i \cdot \sigma_{A_t}^2 / \sqrt{\sigma_{A_t}^2 + \sigma_E^2}\,.$$

In this expression, i is the intensity of selection, h^2, σ_A^2 and σ_P^2 are the heritability, the additive genetic variance and the phenotypic variance of the trait, respectively. The subscript t refers to the generation where the selection occurs, and σ_E^2 is the environmental variance assumed to be constant over time. With the "VCF" method, the evolution of the genetic variance is predicted generation by generation, taking into account both the direct effect of selection on the between family genetic variance (the so-called Bulmer [7] effect) and the combined effects of drift and selection through inbreeding (the so-called Robertson [23] effect); see [27] for details. The global genetic gain was computed as the slope of the linear regression of the cumulated response to selection over generations. The "WB" method provides a prediction of the response to selection in an equilibrium situation, only considering the Bulmer effect. Bulmer-equilibrium genetic parameters were obtained by iteration and subsequently selection response was predicted using equilibrium parameters in the above equation.

Both methods also differ in the prediction of the evolution of the average coefficient of inbreeding. In the "VCF" method, this evolution is predicted generation by generation, on the basis of the probabilities of co-selection of full sibs, paternal half-sibs, maternal half-sibs and candidates which are not sibs, conditionally to the correlation of their values for the selection criterion. Thus VCF considers one generation of pedigree in the prediction of the rate of inbreeding (*i.e.* increased probabilities of co-selection through common sires and dams only, not through grandparents and more distant ancestors). Equation (2) provided a predicted global rate of inbreeding from predicted values of the average coefficient of inbreeding generation by generation.

The "WB" method provides a prediction of the overall rate of inbreeding (ΔF) based on the concept of long-term genetic contributions [31, 32].

The long-term genetic contribution (r_k) of an ancestor k, born in generation t_1, is the proportion of genes in animals born in generation t_2 deriving by descent from k, when $(t_2 - t_1) \rightarrow \infty$. Assuming random mating, the expected rate of inbreeding per generation (ΔF) is proportional to the sum of squared contributions [32]:

$$E(\Delta F) = \frac{1}{4}\sum_{k=1}^{n} r_k^2 \qquad (2)$$

where r_k is the contribution of ancestor k and n is the total number of parents in a generation. In the "WB" method, the long-term genetic contributions of ancestors are predicted conditionally on their breeding value and then, the rate of inbreeding is derived from these predicted long-term contributions [3]. The predicted overall rate of inbreeding enabled values of the average coefficient of inbreeding at each generation to be predicted through equation (1).

A pseudo-prediction of inbreeding generation by generation was also performed by putting the observed genetic contributions of ancestors from successive generations into equation (2). The observed genetic contributions were computed from the pedigree relationships using the probability of the gene origin approach [4]. Then, at each generation, a pseudo-predicted rate of inbreeding was calculated as proportional to the sum of squared-observed contributions to the reference population (generation 8 for the French experiment or 18 for the Dutch experiment) of individuals in the current generation.

3. RESULTS

3.1. Phenotypic trend and response to selection

Figure 1 shows the evolution of the phenotypic mean of each trait both in the line where it was selected and in the control line in the French experiment. The antibody response (ND3) was the trait showing the clearest and the most regular increase in mean under selection (in line 1). Conversely, the increase in the mean of the phagocytose activity (PHA) was low in line 2, and the mean of the cell-mediated immune response (CC) in line 3 showed the largest variation from one generation to the other. In each selected line, the respective means of the two traits which were not selected fluctuated without a significant trend (results not shown). Figure 2 shows the evolution of the phenotypic mean of antibody titre (Ab) in the three Dutch experimental lines. The three lines followed the same fluctuations but significantly differed in their mean from generation 1

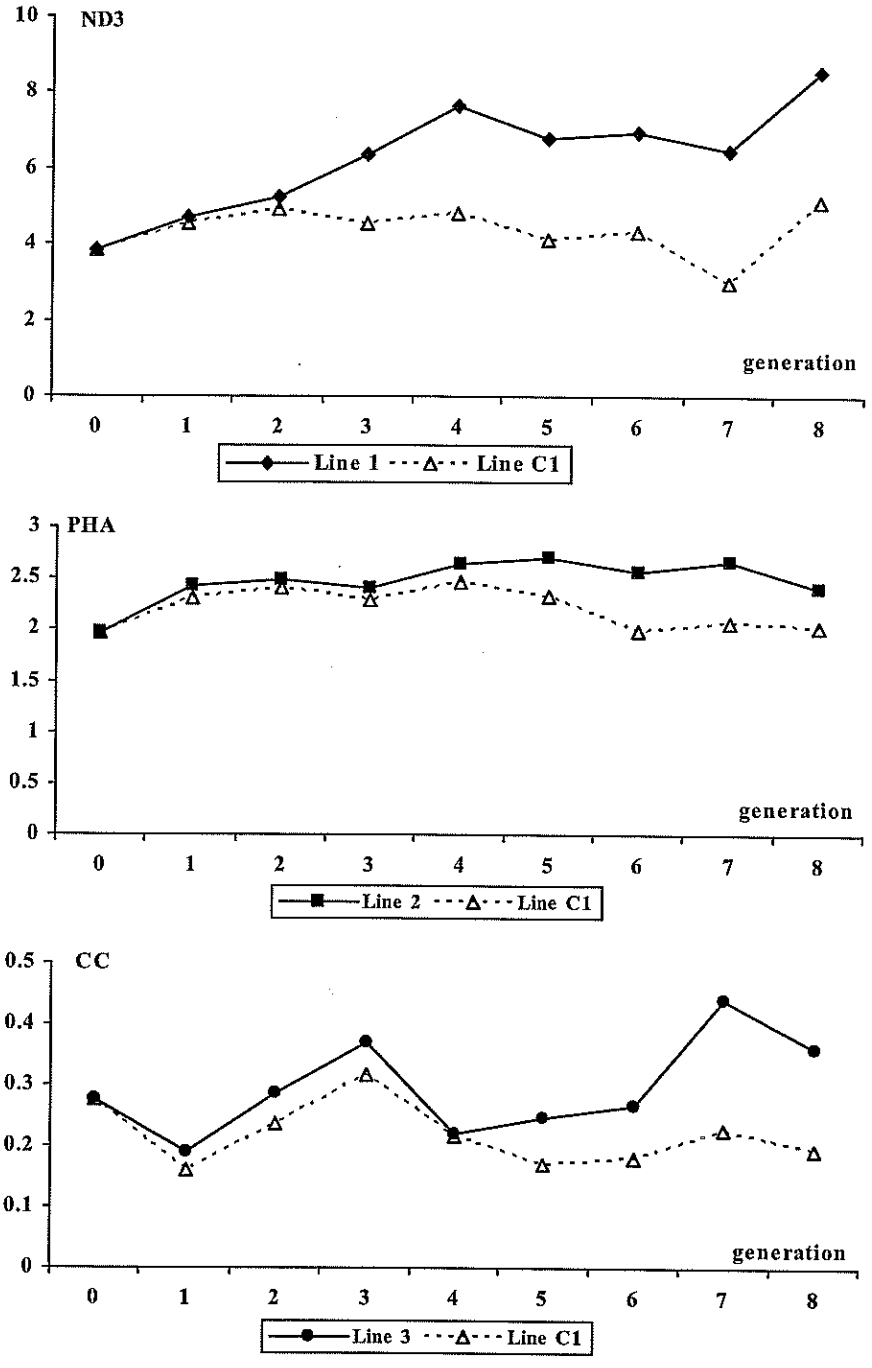


Figure 1. Evolution of the phenotypic mean of each trait in the line where it was selected and in the control line from generations 0 through 8 in the French experiment.

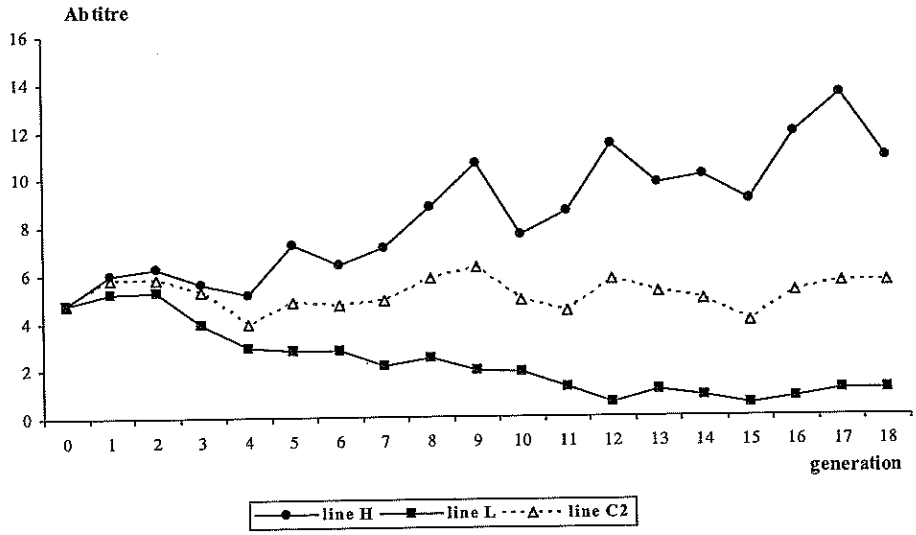


Figure 2. Evolution of the phenotypic mean of antibody (Ab) titre in the three Dutch lines.

on for lines H and L and lines L and C2, and from generation 2 on for lines H and C2.

In the French experiment, realised heritability was found to be 0.24 for line 1, 0.02 for line 2 and 0.16 for line 3. In the Dutch experiment, realised heritability was found to be 0.13 for line H and 0.20 for line L.

Figure 3 shows the evolution of the observed and predicted cumulated responses for the Dutch lines only (the French lines show a similar figure, results not shown). The curve of the observed cumulated response fluctuated: these fluctuations were not taken into account by predictions (the curves of the predicted response were almost linear). The curve of the observed response to selection in line L was much more irregular and slowed down for the last five generations. The curve of predicted response with the "VCF" method was steeper than the one with the "WB" method for the first generations, whereas it was the opposite in the long term.

Table I shows observation and predictions with the two methods of the global genetic gain, for the French and the Dutch experiments. Predictions with the "WB" method are higher than those with the "VCF" method. Both predictions fit the observed values in the Dutch line H and in the French lines 2 and 3. In the Dutch line L, the evolution of the cumulated genetic gain was not regular. However, by culling the stabilised part of the curve (generations 13 through 18), the observed global genetic gain was −0.368 and predictions

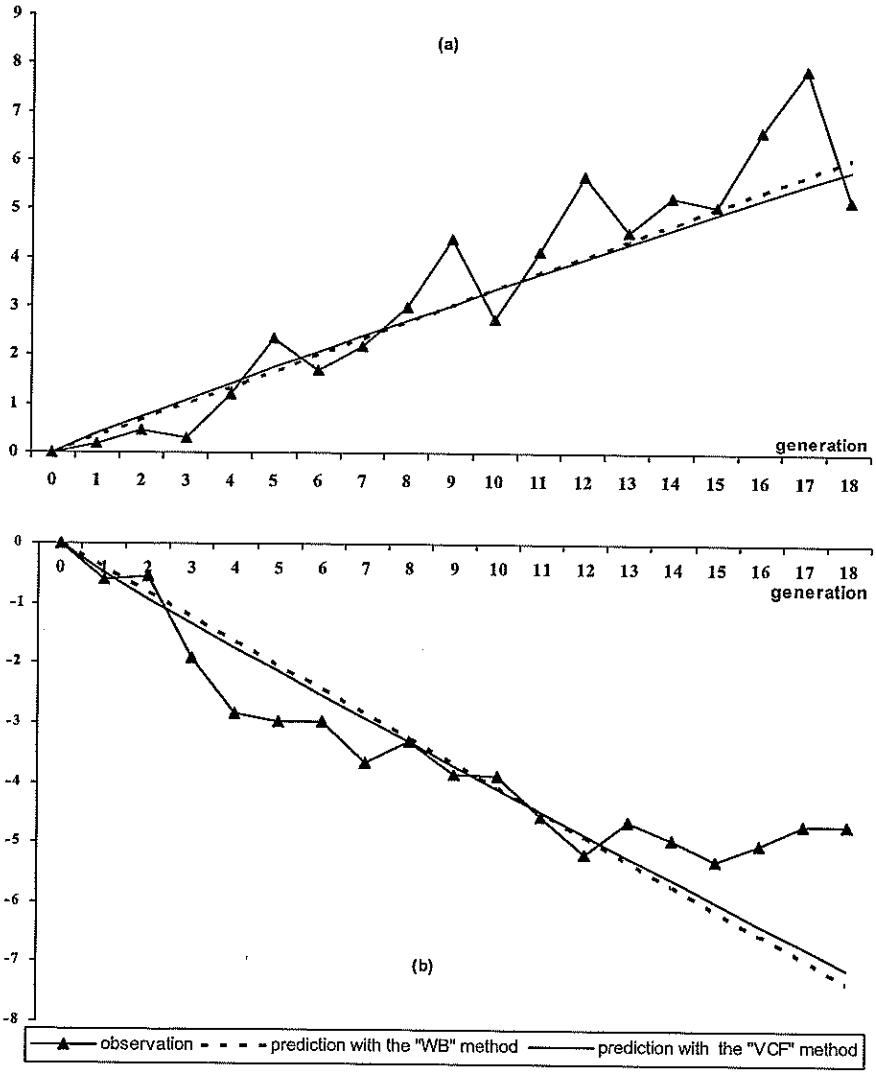


Figure 3. Evolution of the observed cumulated response to selection and the corresponding responses predicted by the two deterministic methods compared for the Dutch selected lines: line H (a) and line L (b).

were closer to the observed value with relative values of 111 for prediction with the "WB" method and 108 for prediction with the "VCF" method. In the French line 1, the observed genetic gain was overestimated by both methods of prediction.

Table I. Observation and predictions of the global genetic gain for the French and Dutch experiments: absolute values and relative values (*in italics*), considering the observed value as the reference (basis = 100).

Global genetic gain	French experiment						Dutch experiment			
	Line 1		Line 2		Line 3		Line H		Line L	
Observations	0.483	*100*	0.070	*100*	0.0220	*100*	0.395	*100*	−0.268	*100*
Predictions										
method WB	0.643	*133*	0.079	*113*	0.0210	*96*	0.336	*97*	−0.407	*152*
method VCF	0.640	*132*	0.076	*109*	0.0209	*95*	0.321	*98*	−0.390	*146*

Table II. Observed inbreeding (%) in the French and Dutch experiments.

Inbreeding (%)	French experiment				Dutch experiment		
	Line 1	Line 2	Line 3	Line C1	Line H	Line L	Line C2
Generation 8	11.48	10.82	10.30	9.90	7.50	8.19	3.34
Generation 18	...	...	...	...	17.77	16.29	7.26

3.2. Inbreeding

Table II shows observed inbreeding for the French and the Dutch experiments. Inbreeding almost identically increased in the four French lines. From generations 0 through 4, the three Dutch lines followed the same trend; however, from generations 5 through 8, the observed inbreeding still increased for lines H and L but stabilised in the control line.

Because there was a higher number of sires and dams at each generation for the Dutch lines than for the French lines, the observed global rate of inbreeding from generations 0 through 8 was lower for the Dutch control line (0.41%) than for the French one (1.26%), according to the Wright equation [33]. Due to true mass selection that occurred in the Dutch selected lines and a higher number of animals in the control line, the observed global rate of inbreeding from generations 0 through 8 was two times higher for lines H and L (0.97% and 1.06%, respectively) than for line C2. However, an increase of the observed global rate of inbreeding in the French experiment was partly offset by the attempt to balance family sizes after that mass selection occurred: the observed global rate of inbreeding from generations 0 through 8 was only 1.47% for line 1, 1.38% for line 2 and 1.31% for line 3.

Figure 4 shows the evolution of the observed, pseudo-predicted and predicted average coefficient of inbreeding for the Dutch selected lines only. Pseudo-predicted inbreeding was higher than predicted inbreeding with the two methods. Prediction with the "WB" method was higher than prediction with the "VCF" method.

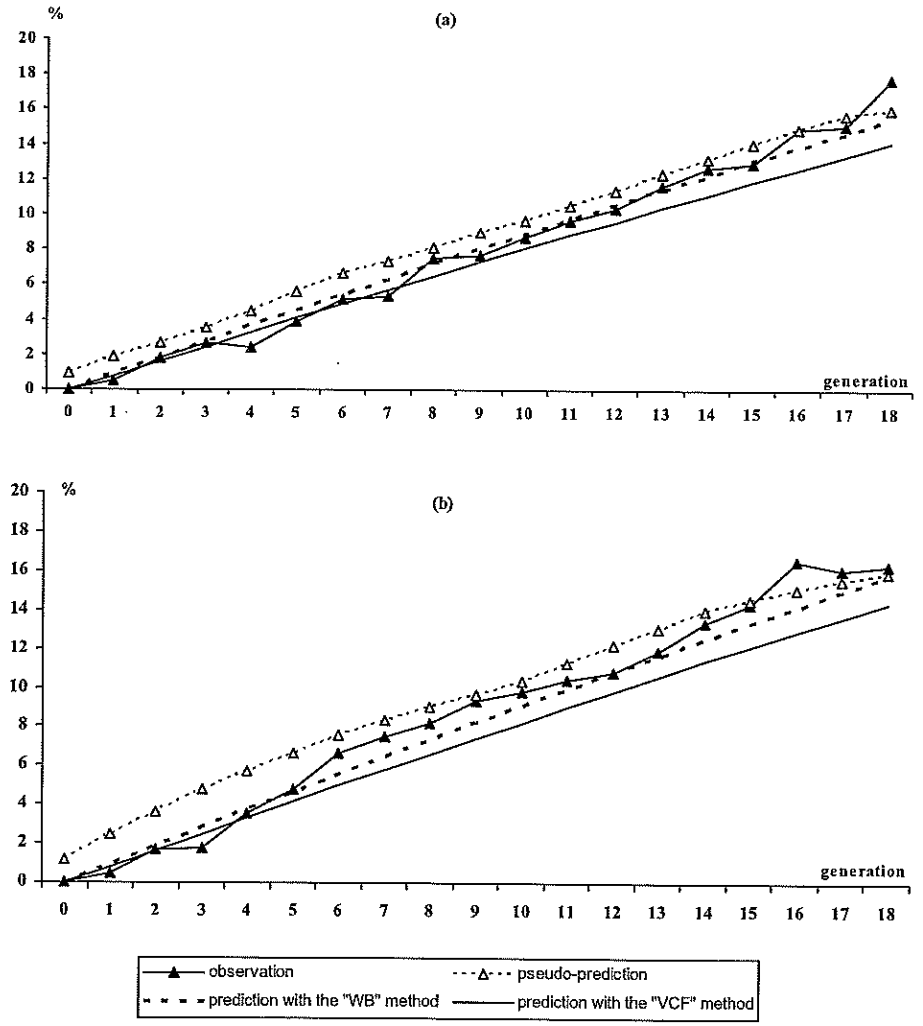


Figure 4. Evolution of observed, pseudo-predicted and predicted inbreeding for the Dutch selected lines: line H (a) and line L (b).

Table III shows observation, pseudo-prediction and predictions with the two methods of the global rate of inbreeding, for the French and Dutch experiments. Pseudo-predictions of the global rate of inbreeding are close to (except for line 2) the observed values on a short-term period (generations 0 through 8) in the French experiment as well as in the Dutch experiment, and underestimated observed values for a longer period (generations 0 through 18). Prediction with the "WB" method over-estimated the global rate of inbreeding in the French experiment whereas they were close to the observed values in

Table III. Observation, pseudo-prediction and predictions of the global rate of inbreeding (%) for the French and the Dutch experiments: absolute values and relative values (*in italics*), considering the observed value as the reference (basis = 100).

Global rate of inbreeding	Generations	French experiment								Dutch experiment								
		Line 1		Line 2		Line 3		Line C1		Line H			Line L			Line C2		
Observations	0 through 8	1.47	*100*	1.38	*100*	1.31	*100*	1.26	*100*	0.97	*100*[a]		1.06	*100*[a]		0.41	*100*[a]	
	0 through 18	...		...		...		...		1.08		*100*[b]	0.98		*100*[b]	0.42		*100*[b]
Pseudo-	0 through 8	1.42	*97*	1.19	*86*	1.27	*97*	1.36	*106*	0.93	*96*[a]		1.03	*97*[a]		0.41	*100*[a]	
predictions	0 through 18	...		...		...		...		0.91		*84*[b]	0.89		*91*[b]	0.39		*93*[b]
Predictions	method WB	1.62	*110*	1.42	*103*	1.45	*111*	1.25*	*99*	0.93	*96*[a]	*86*[b]	0.95	*90*[a]		0.49*	*120*[a]	*117*[b]
	method VCF	1.42	*97*	1.32	*96*	1.34	*102*			0.84	*87*[a]	*78*[b]	0.86	*81*[a]	*88*[b]			

* Random selection: prediction follows the Wright equation *i.e.* $1/(8N_m)+1/(8N_f)$.

[a] Comparison for short-term period (generations 0 through 8).

[b] Comparison for long-term period (generations 0 through 18).

the Dutch experiment. On the contrary, prediction with the "VCF" method underestimated the global rate of inbreeding in the Dutch experiment whereas it was close to the observed values in the French experiment.

4. DISCUSSION

For both French and Dutch lines, individuals were kept and grown in the same "environment" so that they were affected by a common non-genetic year effect. However, a change in the environmental conditions occurred in the French experiment between generations 4 and 5 (use of collective cages instead of individual cages): the observed genetic mean may have been affected by this change. Thus, the observed genetic gain calculated from the evolution of the cumulated genetic gain in the French experiment has to be considered carefully.

Stabilisation of the observed genetic mean for line L could be explained by accessing a selection limit. Indeed, the aim of selection in line L was to decrease the Ab titre. The objective was reached since the phenotypic values of individuals from generations 14 through 18 were close to zero (0.5 at generation 15) and the Ab titre could not be negative.

No distinction was made between males and females in the Dutch experiment whereas a difference of Ab titre exists: females have a higher Ab titre against SRBC than males [20] therefore, the trend in Figure 1 was an average over sexes.

The two methods that were used in this study need to predict the genetic variance over time in order to give an accurate prediction of the cumulated genetic gain. Both methods seem to agree with calculations from the Dutch dataset, although the second method seems to be more accurate. However, predictions and observations were calculated assuming the additive infinitesimal model, implying that the fixation of alleles directly due to selection is ignored and the use of the infinitesimal model may be questionned for two reasons. First, the response to selection is model dependant and the additive infinitesimal model does not take into account changes in additive genetic variance due to selection [19]. Second, MHC genes are supposed to be major genes. Indeed, MHC plays an important role in the immune response and several studies have summarised the association of chicken MHC with resistance to disease [1, 8, 14].

The predictions presented in this paper were obtained using estimated phenotypic variances and heritabilities from the same data as used in the present comparison. Consequently, prediction errors in the estimated heritability and

phenotypic variance may be correlated to the difference between results from this experiment and average results that would have been obtained if the selection experiment had been replicated many times. Ideally, the estimation of variance components and of the "observed" response would have been based on independent data. However, long-term selection experiments in livestock are scarce and we had no independent data available for this purpose.

Pseudo-predicted inbreeding at generation t is defined as being proportional to the sum of squared contributions of individuals in generation t down to generation 0. After several generations, genetic contributions of ancestors stabilise and long-term contributions are reached [32]. Therefore, contributions of individuals from the younger generations have not converged yet. Because drift would increase the loss of diversity from one generation to another, pseudo-predicted inbreeding of these generations should be higher than predictions based on long-term stabilised contributions. Thus pseudo-prediction is higher than prediction with the "WB" method.

Moreover, the pseudo-predicted rate of inbreeding was over-estimated for the oldest generations. Indeed, there is no random mating due to, on the one hand, rules defined in the selection scheme (*i.e.* avoiding full and half-sib mating, no selfing, two sexes). On the other hand, random fluctuation of the mating system may cause deviations from Hardy-Weinberg equilibrium and therefore affect the rate of inbreeding, so that estimation should be based on the following equation:

$$E(\Delta F) = 1/4 \cdot \sum_{k=1}^{n} (1 - \alpha_{I,k}) r_k^2$$

where $\alpha_{I,k}$ is the deviation from Hardy-Weinberg equilibrium in individual k [24]. When mating is not at random, an additional factor reduces heterozygosity: $\alpha_{I,k}$ may be seen as the correlation between alleles within individual k due to mating not at random. Hence, a deviation from random mating has a higher impact on ΔF when the individual in which it occurs has a higher long-term contribution.

The realised contributions of the founders could be considered as "real" long-term genetic contributions because they are calculated over a sufficiently long period to assure convergence. Thus, the contributions of founders calculated from the pedigree do not change anymore after approximately 10 generations. However, drift and inbreeding continue beyond generation 10, indicating that the contributions of founder *alleles* must fluctuate, despite convergence of founder contributions when calculated from the pedigree. The reason is that the contributions calculated from the pedigree are 'expected contributions given the pedigree'; truly realised contributions may deviate from pedigree

contributions due to Mendelian sampling, which is not observable from the pedigree. Thus realised contributions of founders and long-term contributions calculated from the pedigree are not strictly equal. This discrepancy is accounted for in the derivation of the "WB" method, and predictions from it should therefore fit pseudo-predictions for the oldest generations.

5. CONCLUSION

This study confirmed, on an experimental basis, that modelling is an efficient approach to make useful predictions of the evolution of selected populations, despite that basic assumptions considered in the models (polygenic additive model, normality of the distribution, base population at the equilibrium, etc.) are not met in reality. The two deterministic methods yielded results that were close to real data, especially when the selection scheme followed the rules of strict mass selection.

REFERENCES

[1] Bacon L.D., Influence of the MHC on disease and productivity, Poult. Sci. 66 (1997) 802–811.

[2] Bijma P., Woolliams J.A., Prediction of rates of inbreeding in populations selected on best linear unbiased prediction of breeding value, Genetics 156 (2000) 361–373.

[3] Bijma P., van Arendonk J.A.M., Woolliams J.A., A general procedure to predict rates of inbreeding in populations undergoing mass selection, Genetics 154 (2000) 1865–1877.

[4] Boichard D., Maignel L., Verrier E., The value of using probabilities of gene origin to measure genetic variability in a population, Genet. Sel. Evol. 29 (1997) 5–23.

[5] Bordas A., Tixier-Boichard M., Mérat P., Direct and correlated responses to divergent selection for residual food intake in Rhode Island Red laying hens, Br. Poult. Sci. 33 (1992) 741–754.

[6] Bovenhuis H., Bralten H., Nieuwland M.G.B., Parmentier H.K., Genetic parameters for antibody response of chickens to sheep red blood cells on a selection experiment, Poult. Sci. 81 (2002) 309–315.

[7] Bulmer M.G., The effect of selection on genetic variability, Am. Nat. 105 (1971) 201–211.

[8] Bumstead N., Millard B., Barrow P., Cook J.K.A., Genetic basis of disease resistance in chickens, in: Owen J.B., Axford R.F.E. (Eds.), Breeding for disease resistance in farm animals, CAB International, Melksham, UK, 1991, pp. 10–23.

[9] Caballero A., Developments in prediction of effective population size, Heredity 73 (1994) 657–679.

[10] Clark A.G., Szumski F.M., Bell K.A., Keith L.E., Houtz S., Merriwether D.A., Direct and correlated responses to artificial selection on lipid and glycogen contents in *Drosophila melanogaster*, Genet. Res., 56 (1990) 49–56.
[11] Colleau J.J., Moureaux S., Briend M., Bechu J., A method for the dynamic management of genetic variability in dairy cattle, Genet. Sel. Evol. 36 (2004) 373–394.
[12] Falconer D.S., Mackay T.F.C., Introduction to quantitative genetics, 3rd edn., Longman Scientific & Technical, Essex, UK, 1996.
[13] Huby M., Griffon L., Moureaux S., de Rochambeau H., Danchin-Burge C., Verrier E., Genetic variability of six French meat sheep breeds in relation to their genetic management, Genet. Sel. Evol. 35 (2003) 637–655.
[14] Kaufman J.F., Lamont S.J., The chicken major histocompatibility complex, in: Schook L.B., Lamont S.J. (Eds.), The major histocompatibility complex in domestic animal species, CRC Press, Boca Raton, Florida, 1996, pp. 35–64.
[15] Laval G., Chevalet C., Pitel F., Amigues Y., Bordas A., Coville J.L., Tixier-Boichard M., Modélisation et évaluation expérimentale de la diversité neutre ou adaptative dans le cadre de lignées sélectionnées chez la poule, in: Proceedings of 4e Colloque national du BRG, 14–16 octobre 2002, La Châtre, France, Ed. BRG, Paris.
[16] Martinez V., Bunger L., Hill WG., Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal model assumptions, Genet. Sel. Evol. 32 (2000) 3–21.
[17] Meuwissen T.H.E., Maximizing the response of selection with a predefined rate of inbreeding, J. Anim. Sci. 75 (1997) 934–940.
[18] Moureaux S., Verrier E., Ricard A., Mériaux J.C., Genetic variability within French race and riding-horse breeds from genealogical data and blood marker polymorphism, Genet. Sel. Evol. 28 (1996) 83–102.
[19] Ollivier L., On the use of animal models in the analysis of selection experiments, Genet. Sel. Evol. 31 (1999) 135–148.
[20] Parmentier H.K., Niewland M.G., Rijke E., De Vries Reilingh G., Schrama J.W., Different response to vaccines and divergent body weight of chicken lines selected for high and low humoral responsiveness to sheep red blood cells, Avian. Dis. 40 (1996) 634–644.
[21] Pinard M.-H., van Arendonk J.A.M., Nieuland M.G.B, van der Zipp A.J., Divergent selection for immune responsiveness in chickens: estimation of realized heritability with an animal model, J. Anim. Sci. 70 (1992) 2986–2993.
[22] Pinard-van der Laan M.-H., Immune modulation: the genetic approach, Vet. Immunol. Immunopathol. 87 (2002) 199–205.
[23] Robertson A., Inbreeding in selection programs, Genet. Res. 2 (1961) 189–194.
[24] Sanchez L., Woolliams J.A., Impact of nonrandom mating on genetic variance and gene flow in populations with mass selection, Genetics 166 (2004) 527–535.
[25] Santiago E., Caballero A., Effective size of populations under selection, Genetics 139 (1995) 1013–1030.
[26] Van Raden P.M., Accounting for inbreeding and crossbreeding in genetic evaluation for large populations, J. Dairy Sci. 75 (1992) 3136–3144.

[27] 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.

[28] Verrier E., Colleau J.J., Foulley J.L., Long term effects of selection based on the animal model BLUP in a finite population, Theor. Appl. Genet. 87 (1993) 446–454.

[29] Woolliams J.A., Bijma P., Predicted rated of inbreeding in population undergoing selection, Genetics 154 (2000) 1851–1864.

[30] Woolliams J.A., Mäntysaari E.A., Genetic contributions of Finnish Ayrshire bulls over four generations, Anim. Sci. 61 (1995) 177–187.

[31] Woolliams J.A., Wray N.R., Thompson R., Prediction of long-term contributions and inbreeding in populations undergoing mass selection, Genet. Res. Camb. 62 (1993) 231–242.

[32] Wray N.R., Thompson R., Prediction of rates of inbreeding in selected populations, Genet. Res. Camb. 55 (1990) 41–54.

[33] Wright S., Evolution and the genetics of population, in: The theory of gene frequencies, Vol. 2, Univ. of Chicago, Chicago, 1969.

II. A comparison of the evolution of some parameters based on pedigree analysis

The effect of selection strategies on the evolution of the within-population variability assessed by pedigree data has often been investigated. As an example, among others, Huby *et al.* (2003) compared the within-population variability of some French meat sheep breeds on the light of their genetic management. Another interesting example was the case of the dairy sheep breed Lacaune (Palhière *et al.*, 2002): the breed was split into two separate selection schemes that have been independently conducted for the last twenty years. Both schemes paid attention to management of genetic diversity but one of them applied rules in a more constant way across years. As a consequence, the observed inbreeding rate was lower for the first scheme and higher in the second one. Yet objectives of both schemes were improvement of milk quantity and milk quality, and both started from the same large population, and both achieved the same genetic gains during the period analysed. It was concluded that procedures aiming at preserving the genetic variability may meet success not at the expense of genetic gains.

Then, the present study aims at analysing effect of selection scheme on inbreeding and also other criteria of genetic variability based on pedigree data for selection experiments undergoing different selection scheme, with selected lines and control line diverging from the same population.

1. METHODS

Individuals inbreeding coefficients, number of founders, effective number of founders and effective numbers of ancestors were computed from the pedigree by using the PEDIG package (ref. http://www-sgqa.jouy.inra.fr/diffusions/pedig/pedig.htm).

The programme computed the probabilities of gene origin for a reference population. In order to obtain the three parameters based on probabilities of gene origin (Boichard *et al.*, 1997) at each generation, calculation was done by successively deleting the most recent generation from the pedigree, *i.e.* successively changing the reference population. No distinction was made between males and females.

For details on these methods, see chapter 1.

2. RESULTS AND DISCUSSION

Figure 2 shows number of founders, effective number of founders and effective number of ancestors over generations for Line 1. The evolution of these parameters was similar in other lines (results not shown).

Total number and effective number of founders rapidly decreased and stabilized in generation 2 and generation 1, respectively. Effective number of ancestors constantly decreased. In generation 8 in Line 1, number of founders, effective number of founders and effective number of ancestors was in average over the lines 33 (±1.7), 20 (±1.2) and 12 (±1.3), respectively.

Evolution of number of founders, effective number of founders and effective number of ancestors over generations were equivalent in the two Dutch lines but values of different parameters were higher for the control line. Figure 3 shows number of founders, effective number of founders and effective number of ancestors over generations for Line H (a) and for Line C2 (b).

Number and effective number of founders rapidly decreased and stabilized in generation 3 and generation 2, respectively. Effective number of ancestors constantly decreased. In generation 8, number of founders, effective number of founders and effective number of ancestors was in average over the selected lines 51 (±1.4), 24 (±4.1) and 18 (±2.9), respectively and were 71, 42 and 40 respectively for the control line.

Starting from a base population, the reduction of the total number and the effective number of founders is rapid by nature, as illustrated *via* simulations by Verrier *et al*. (1994): a large part of the initial lineages are lost when the selection process starts and, due to the increase of relationship between animals, the representation of the remaining lineages is stabilized after a small number of generations. The reduction of the effective number of ancestors occurs very early too but, contrary to the effective number of founders, it does not stabilize, due to possible bottlenecks during the successive generations (Boichard *et al.*, 1997).

After stabilization, i.e. since creation of lines, effective number of founders of the French selected lines represented two thirds of the total number of founders. This ratio was the same for the random-bred lines (Line C1 and C2) whereas it represented only half for the Dutch selected lines.

Effective number of founders shows whether expected contributions of founders are balanced but does not account for bottlenecks further on in the pedigree. Bottlenecks are taken into

account in effective number of ancestors. It appears that effective number of founders and effective number of ancestors were almost the same for the random-bred Dutch line (Line C2) but the ratio (effective number of ancestors divided by effective number of founders) was smaller when selection occurred (Line H and L). In comparison with a selection procedure aiming to balance the family representation at each generation, mass selection tends to favour bottlenecks. Ratio of the French control line was lower than ratio of the Dutch control line but the number of parents selected was larger in the Dutch selection scheme than in the French one: loss of diversity due to bottlenecks increases with selection intensity.
Effective number of ancestors is concerned with the whole pedigree since major ancestors may be founders or not: an individual from any intermediate generations (generation 1 through 7) who affects the current generation is taken into account. Effective number of ancestors reflects effects of the breeding scheme on pedigree and their consequences on gene loss in the population.

Effective size of the population based on inbreeding rate (N_{eI}) can be defined as the number of individuals with random mating and random selection that would have the same rate of inbreeding as the population under study. It is calculated as: $N_{eI} = 1/2\Delta F$ (e.g., Falconer and Mackay, 1996).

Wray and Thompson (1990) have shown that $E(\Delta F) = 1/4\sum_{i=1}^{n} r_i^2$, where r_i is the realized contribution of one ancestor. Afterwards, Wooliams and Bijma (2000) showed that the rate of inbreeding could be expressed in terms of expected contributions (r_k ; see chapter 1) and under hypothesis of random mating, we may write: $E(\Delta F) = 1/4\sum_{i=1}^{n} r_k^2$

As $f_e = \sum_{k=1}^{f} r_k^2$ (see Chp.1), it leads us to the following equality: $N_{eC} = 2\, f_e$. Thereafter, we will refer to N_{eC} as the effective size of population based on contributions; N_{eC} stands here for the effective size of the founder generation. Boichard *et al*. (1997) showed that the computation of inbreeding is more sensitive to lacks in pedigree knowledge than the computation of probabilities of gene origin. In the present study, pedigree is complete, which provides good conditions to compare N_{eI} and N_{eC} (Table 1).
Difference appears because those two parameters do not exactly deal with the same concept. Indeed, N_{eI} is based on inbreeding rate from generation 0 through generation 8, and more precisely, $\Delta F = \Delta F_{(G0\ to\ G1)} + \Delta F_{(G1\ to\ G2)} + ... + \Delta F_{(G7\ to\ G8)}$. Drift that has modified genetic

contributions in generations prior to generation 8 is here taken into account. However, f_e focuses only on individuals of Generation 0 and deals with their future contributions to Generation 8. Drift that will occur in contributions of individuals born in Generation 1 through Generation 8 is not here taken into account. In short, we could say that, on one hand, effective size of the population based on inbreeding rate "looks backwards" by considering influence of generation 0 but also of intermediate generations on the current generation; on the other hand, effective number of the population based on contributions relates to a single generation only and "looks forwards" by considering the future influence of generation 0 on the current generation. It seems that N_{eI} is more accurate for the present study.

The concept of effective number of founders was already criticized by Lacy (1989) for the same reasons since it does not take into account loss of genetic diversity by genetic drift in subsequent generations, which would be relevant in conservation of small populations with potential bottlenecks in the pedigree. Caballero and Toro (2000) also commented that in order to minimize the increase of inbreeding, one should minimize variances of contributions from all generations, not only founders. To overcome the problem, Lacy (1989, 1995) introduced the concept of founder genome equivalents, N_{ge}, which includes genetic drift occurring during the pedigree development and is defined as follows: $N_{ge} = 1 \big/ 2\overline{\Phi}_t$

where $\overline{\Phi}_t$ is the average kinship coefficient at generation t. Originally, N_{ge} was a function of expected proportions of surviving alleles of the founders at generation t:

$$N_{ge} = \frac{1}{\sum_{i=1}^{f} \overline{\Phi}_{i(0,t)} \Big/ r_{i(0,t)}}$$

where $r_{i(0,t)}$ is the expected proportion of surviving alleles at generation t from the i^{th} founder (from generation 0) and $\overline{\Phi}_{i(0,t)}$ is the average kinship coefficient between ancestor i in generation 0 and descendants in generation t. But for practical reasons, the first definition of founder genome equivalents should be preferred since it does not need complex computer calculations along the pedigree and kinship coefficient may be directly derived from the additive relationship matrix (Caballero and Toro, 2000).

Figure 2: Number of founders (black), effective number of founders (grey) and effective number of ancestors (white) over generations for Line 1 from the French experiment.

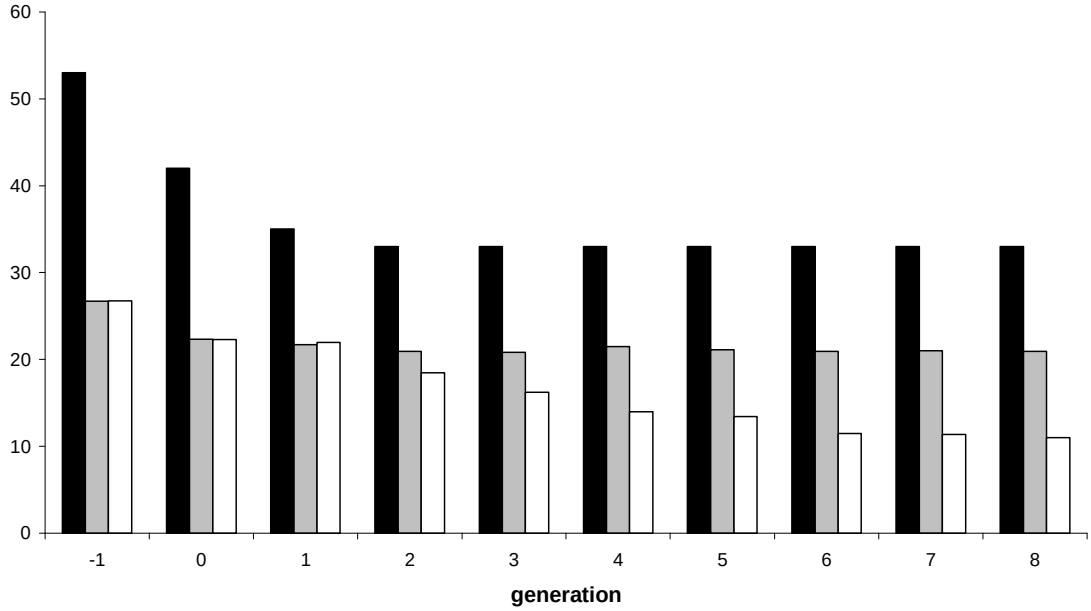
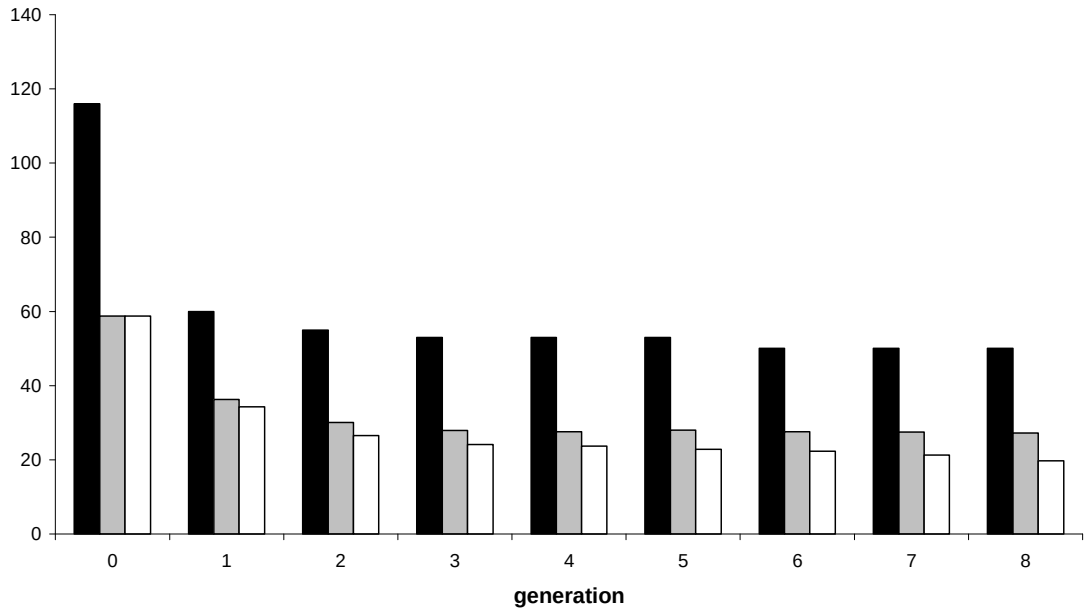


Figure 3: Number of founders (black), effective number of founders (grey) and effective number of ancestors (white) over generations for Line H (*a*) and Line C2 (*b*) from the Dutch experiment.

(*a*)



(*b*)

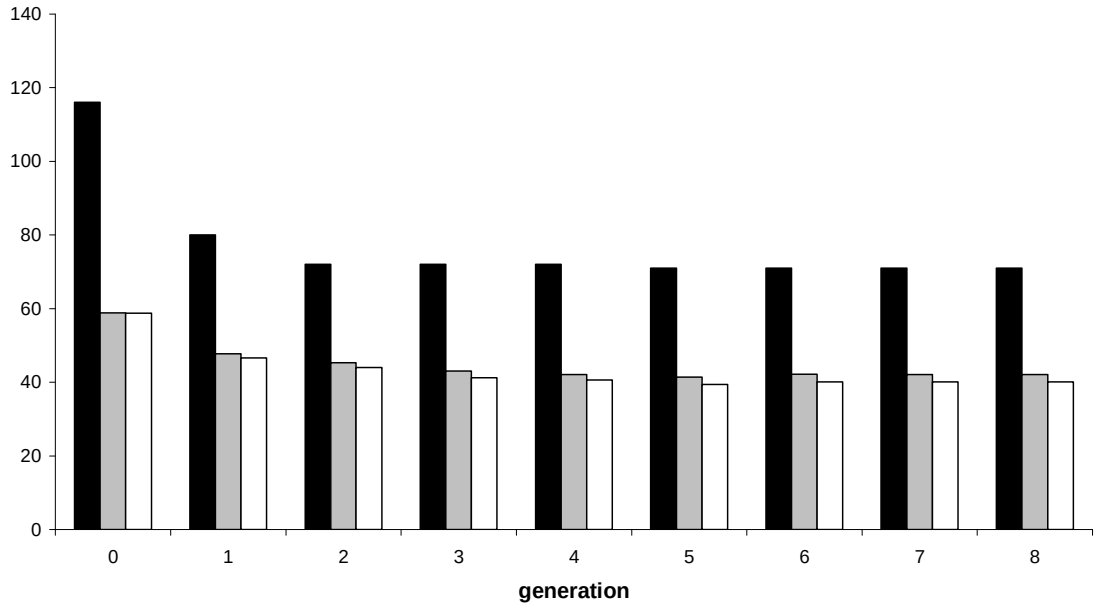


Table 1: Effective size of the population based on inbreeding rate (N_{eI}) and on contributions (N_{eC}) for the French and the Dutch experiments.

		$\boldsymbol{N_{eI}}$	$\boldsymbol{N_{eC}}$
French experiment	Line 1	34	42
	Line 2	36	37
	Line 3	38	38
	Line C1	40	41
Dutch experiment	Line H	52	54
	Line L	47	43
	Line C2	119	84

CHAPTER 3: Estimation and evolution of genetic parameters and polymorphism evolution of a candidate gene

This chapter deals with an analysis of genetic diversity from the French experimental lines and aims to estimate evolution of genetic parameters and polymorphism evolution of a candidate gene.

The first goal of this study was to give actual estimates of the genetic parameters (additive genetic variance, heritability, genetic correlations) in the base population. Second, MHC genes being good candidate genes for the selected traits in immune response, effect of MHC genes on the traits was estimated and analysis of the MHC polymorphism within lines was handled. This evolution will be compared to different theoretical predictions, considering MHC genes to be neutral or not. Third, evolution over time of the genetic diversity may be assessed by estimating the parameters with different samples of the data, in order to investigate the influence of available data on estimations.

I. Actual genetic parameters and effect of a candidate gene

(see article)

Genet. Res., Camb., Page 1 of 7.
doi:10.1017/S0016672306008044 Printed in the United Kingdom

On the need for combining complementary analyses to assess the effect of a candidate gene and the evolution of its polymorphism: the example of the Major Histocompatibility Complex in chicken

VALERIE LOYWYCK*, MARIE-HELENE PINARD-VAN DER LAAN, ISABELLE GOLDRINGER† AND ETIENNE VERRIER
INRA/INA P-G, UMR Génétique et Diversité Animales, 78352 Jouy-en-Josas, France

(*Received 15 August 2005 and in revised form 17 January 2006*)

Summary

The aim of this paper is to combine different but complementary approaches to check the neutrality of a given locus in a selected population. Analysis was undertaken through the polymorphism's evolution compared with that predicted under the effect of drift and through the analysis of the variance components of the measured traits, considering the effect of the locus as either a fixed or a random effect. This study deals with the case of the MHC locus, using both data from experimental lines of chicken selected for three different criteria of immune response, and frequencies of the genotyped haplotypes over time. Both the evolution of the polymorphism and the variance components approach have led to the conclusion that the MHC locus has an effect on the trait affecting antibody production against the Newcastle disease virus. Results have also highlighted the interest in using various methods in the case of low allelic frequencies. However, none of the common hypotheses, overdominance or frequency-dependent selection, was sufficient to explain the observed variation of the MHC polymorphism, which was displayed by the temporal variation of the allelic frequencies.

1. Introduction

The Major Histocompatibility Complex (MHC) plays an important role in the immune system of all vertebrates. In farm animals, knowledge of the MHC genes has been growing and, in several chicken studies, the MHC has been shown to be involved in immune response to pathogens and disease resistance traits (reviewed by Bacon, 1987; Bumstead *et al.*, 1991; Kaufman & Lamont, 1996). Yet, these data are still limited and selection experiments have mainly focused on the observed change in allelic frequencies between selected lines, but not its evolution over time.

The observed evolution over time of the polymorphism at a given locus results mainly from the joint effects of drift and selection, which are known to be two main factors affecting evolution of the genetic variability within a closed population. In order to check the neutrality of a given locus, results on its polymorphism's evolution should be compared with the expected results under the assumption of pure genetic drift. Another way is to analyse the variance components of a measured trait known to be selected at the phenotypic level and to estimate the effects of the different alleles on this trait.

The aim of this paper is to combine these different approaches for the case of the MHC locus, using data from experimental lines of chicken selected for different criteria of immune response.

* Corresponding author. Tel: +33 1 34652171. Fax: +33 1 34652210. e-mail: loywyck@inapg.fr
† Current address: INRA/UPS/CNRS/INA P-G, UMR de Génétique Végétale, Ferme du Moulon, 91190 Gif-sur-Yvette, France.

2. Material

(i) *Selection design*

Four experimental lines of chickens have been developed since 1994 from an unselected base population of White Leghorn chickens (Pinard-van der Laan, 2002). Three of these lines were selected for high values according to three different criteria of immune response: antibody response 3 weeks after vaccination against the Newcastle disease virus (line 1, trait ND3),

cell-mediated immune response at 9 weeks of age (line 2, trait PHA) and phagocytic activity at 12 weeks of age (line 3, trait CC). The three lines underwent mass selection with a restriction on the contribution of the different families (sizes of the different half-sib families were approximately balanced). The fourth line was a control line (line 4), in which the parents were chosen at random. Within each line and at each generation, 15 males and 30 females out of about 100 candidates of each sex were chosen as parents for the next generation. Mating was at random, except that full- and half-sib matings were avoided. This selection programme was conducted for nine discrete generations (generation 1 to generation 9, generation 0 being considered as the base population). Results of the selection response and the evolution of inbreeding were reported by Pinard-van der Laan (2002) and Loywyck *et al.* (2005). A total of 7550 animals have been measured for the three traits, whatever the status of their line (control or selected) and the nature of the selection criterion.

(ii) *MHC typing*

MHC is a complex of three regions comprised of several genes encoding for three classes of molecules: B-F and B-L gene products which are equivalent to class I and class II gene products of mammals, respectively, and B-G gene products which are equivalent to class IV gene products, specific to birds (Miller *et al.*, 2004). From the begining of the selection, all birds were typed for the MHC antigens using alloantisera produced from the lines. Six B-G haplotypes were found: B^{15}, B^{19}, B^{21}, B^{34}, B^{121} and B^{124}. Each haplotype identified by serology corresponded to a unique B-G and B-F restriction fragment length polymorphism (RFLP) pattern, with the exception of the B^{21} and the B^{121} haplotypes which are serologically different and have different B-G RFLP patterns but the same B-F RFLP pattern. Because the different analyses did not reveal any significant difference between the two haplotypes, the results presented will consider five B-F haplotypes only: B^{15}, B^{19}, B^{34}, B^{124} and B^{21} (grouping the B-G^{21} and B-G^{121}). Twenty-one genotypes were observed within the four lines.

3. Methods

(i) *Variance component analysis and estimation of haplotypes effects*

In a first step, variance components of the three immune traits (ND3, PHA and CC) were estimated by the Restricted Maximum Likelihood (REML) method, using VCE software (Groeneveld, 1997). Because previous results obtained from the first six generations showed no genetic correlation between the three traits (Pinard-Van der Laan, 2002), single-trait analyses were performed. Performance and pedigree data from all generations, up to the base population, were used. On the basis of simulation results by Sorensen *et al.* (2003), for a given trait data from both the line selected for this trait and the control line were used.

For each trait, three animal models were compared: the first did not include the effect of the MHC genotype and the other two differed in the way this effect was taken into account:

$$Y_{ijl}=\mu+gener_i+sex_j+A_{ijl}+E_{ijl} \quad \text{(model 1)}$$

$$Y_{ijkl}=\mu+gener_i+sex_j+z_k+A_{ijkl}+E_{ijkl} \quad \text{(model 2)}$$

$$Y_{ijkl}=\mu+gener_i+sex_j+Z_k+A_{ijkl}+E_{ijkl} \quad \text{(model 3)}$$

In these equations, subscripts i, j, k and l refer to the generation number, the sex of the animal, its MHC genotype and the animal itself, respectively; Y is the performance; μ is the overall mean; *gener* and *sex* are the environmental fixed effects of the generation and the sex of animals, respectively; A is the additive genetic value, assumed to be of polygenic origin; and E is the random error. The effect of the MHC genotype is represented by z or Z, this effect being considered as fixed (model 2) or random (model 3).

In a second step, the breeding value (A) of each animal and the random effect (Z) of each MHC genotype were predicted by the BLUP method under model 3, using the genetic parameters estimated during the first step under the same model, and assuming no correlation between the two variables A and Z. PEST software (Groeneveld, 1990) was used to perform this analysis. Due to the presence of rare alleles, only model 3 was used here, because considering the genotype effect as random was the only way to avoid estimation problems for several genotypes with a very small number of available performance data.

In a third step, the effects of the different haplotypes were compared by the method of contrasts, under model 2. This method consists of comparing the mean of the effects of genotypes including one haplotype with the mean of the effects of genotypes including another haplotype, assuming that the genotype effect may be considered as two additive haplotype effects. As the use of the usual statistical test was complicated because the degrees of freedom due to error that are not defined in a mixed-effect model (model 3), only model 2 was used here. Moreover, grouping in the analysis the rare genotypes with less rare ones prevents the estimation problems evoked in the second step. The method was performed using data from all generations over the four lines. First, the effects of each haplotype on the three traits were estimated; second, the effects of homozygous

Table 1. *Estimation of the genetic parameters for each trait as a percentage of the total phenotypic variation (SE, in percentage), according to the three different models considered: with (fixed or random) or without an MHC effect*

Trait	Parameter	Model 1: without MHC effect	Model 2: MHC as a fixed effect	Model 3: MHC as a random effect
ND3 [line 1]	σ^2_A	33·4 (1·1)	32·1 (1·0)	31·4 (1·1)
	σ^2_{MHC}	–	–	2·3* (0·3)
	σ^2_E	66·6 (0·8)	67·9 (0·7)	66·2 (0·8)
PHA [line 2]	σ^2_A	11·9 (2·0)	11·9 (1·9)	11·9 (1·8)
	σ^2_{MHC}	–	–	0·1 NS (0·1)
	σ^2_E	88·1 (2·2)	88·1 (2·2)	88·0 (2·1)
CC [line 3]	σ^2_A	23·7 (13·0)	23·1 (13·1)	22·9 (13·0)
	σ^2_{MHC}	–	–	0·8 NS (2·3)
	σ^2_E	76·3 (11·4)	76·9 (12·3)	76·3 (12·2)

genotypes were compared with the effects of heterozygous genotypes in order to test the underlying hypothesis of additivity.

(ii) *Evolution of haplotype frequencies*

Within each line and at each generation, the haplotype frequencies were calculated. Thus, a possible effect of drift was tested. First, the inbreeding effective size (N_{eI}) was estimated from the observed rate of inbreeding (ΔF), according to the classical formula:

$$\hat{N}e_I = \frac{1}{2 \cdot \Delta F}$$

The rate of inbreeding was computed at each generation from the observed values of the average coefficient of inbreeding computed from pedigree data (Loywyck *et al.*, 2005). Next, the comparison between observed and expected change in haplotype frequencies was performed in two complementary ways.

First, a 95% confidence interval of the frequency was determined for each haplotype separately in each generation and in each line. Populations with discrete generations, random choice of parents and random mating were simulated. Because generations did not overlap and due to the stability of the management rules which involved a balance between families, the pure drift situation was simulated using the concept of effective population size. For a given line and a given haplotype, a population was simulated with an effective size equal to the estimated value of N_{eI} and with an initial frequency equal to the observed initial value. In each generation, the bounds of the 95% confidence interval were empirically determined on the basis of the frequencies observed in 5000 independent replicates.

Second, the MHC locus was considered as a whole: an estimate of the standardized temporal variance in allelic frequency, f (Waples, 1989), was computed for each line over the nine generations; the f_c estimator of f, proposed by Nei & Tajima (1981), was used:

$$\hat{f}_c = \frac{1}{k} \sum_{i=1}^{k} \frac{[x_i - y_i]^2}{\frac{x_i + y_i}{2} - x_i \cdot y_i}$$

where k is the number of segregating alleles, x_i is the frequency of allele i at generation 0 and y_i the frequency of this allele at generation t. The estimate of the variance effective size (N_{eV}) of each selected line was directly deduced from the value of f_c, using the equation of Nei & Tajima (1981), as there is no sampling variance because the exact values for frequencies were available (all individuals at each generation had been genotyped):

$$\hat{N}e_V = \frac{t}{2\hat{f}_c}.$$

This value was compared with the value of the inbreeding effective size (N_{eI}). Moreover, as in Goldringer & Bataillon (2004), the observed value of f_c was compared with the distribution of f_c obtained from a series of simulations of populations with the same initial allelic frequencies and the same inbreeding effective size.

4. Results

(i) *Variance components analysis*

Table 1 shows the estimates of the genetic parameters for each trait, according to the three different models considered. When the MHC effect was not taken into account (model 1), the estimated value of the heritability was 0·33, 0·12 and 0·24 for traits ND3, PHA and CC, respectively. These values for traits ND3 and PHA are consistent with those reported

Table 2. *Estimation of the effects of the MHC haplotypes on the three traits, using the method of contrasts and considering the MHC effect as fixed*

Haplotype	Trait PHA		
15	$0{\cdot}175^{ab}$	$0{\cdot}049^{a}$	$0{\cdot}008^{bc}$
19	0^{b}	$0{\cdot}008^{a}$	$0{\cdot}006^{bc}$
21	$0{\cdot}449^{a}$	0^{a}	$0{\cdot}010^{ab}$
34	$0{\cdot}441^{a}$	$0{\cdot}050^{a}$	$0{\cdot}033^{a}$
124	$0{\cdot}356^{a}$	$0{\cdot}047^{a}$	0^{c}

Values are given considering the lowest effect as the reference, e.g. values of the effects of haplotypes on trait ND3 are given as a comparison with the value of the effect of B^{19} haplotype.
^{abc} Two haplotypes sharing the same letter have effects that are not significantly different.

by Pinard-van der Laan (2002) – 0·35 and 0·1 3, respectively – but the estimate for trait CC was higher than that of Pinard-van der Laan (0·15). In comparison with model (1), taking into account the MHC effect led to changes in estimated values for trait ND3 only: for ND3, when the MHC effect was considered as fixed, the estimated value of the additive variance (σ^2_A) was reduced by 5·9 %; when the MHC effect was considered as random, the variance of this effect was found to be significantly different from zero and represented 6·9 % of the total genetic variance (including both the polygenic and the MHC components).

(ii) *Effect of the MHC haplotypes and comparison between heterozygotes and homozygotes genotypes*

Table 2 shows the estimates of the effects of the MHC haplotypes using the method of contrasts. Compared with B^{21}, B^{34} and B^{124}, the B^{19} haplotype had a significant negative effect on trait ND3 and, compared with B^{21} and B^{34}, the B^{124} haplotype had a significant negative effect on trait CC. No significative effect of any haplotype was observed for trait PHA.

In addition, significantly ($P<0{\cdot}005$) lower values for trait CC were associated with heterozygote genotypes than with the homozygote genotypes. Moreover, there was a significantly ($P<0{\cdot}003$) lower value for trait ND3 associated with the heterozygote genotype B^{15}–B^{19} than with the corresponding homozygote genotypes B^{15}–B^{15} and B^{19}–B^{19}, and a significantly ($P<0{\cdot}008$) lower value for trait CC associated with the heterozygote genotype B^{19}–B^{124} than with the corresponding homozygote genotypes B^{19}–B^{19} and B^{124}–B^{124}.

(iii) *Evolution of haplotype frequencies*

Fig. 1 presents the evolution of the observed frequencies of the haplotypes within each selected line and within the control line. The B^{19} haplotype was lost within line 1 at generation 6, and the B^{34} haplotype was lost within line 2 at generation 7 and within line 3 at generation 5. Only the evolution of the frequency of the B^{19} haplotype within line 1 exceeded the 95 % confidence interval under the asumption of drift. For all the other haplotypes in all four lines, no deviation was observed from the confidence interval (for the sake of clarity in the figures, this interval was not shown in these cases).

Table 3. *Estimated effective sizes of the population (for each line over the nine generations) based on the temporal variation approach N_{eV}, and on the pedigree approach, N_{eI}*

Line	f_c at MHC locus (P value)	N_{eI}	N_{eV} [95 % CI]
1	0·0884 (0·599)	34	51 [6–142]
2	0·0694 (0·663)	36	65 [8–181]
3	0·1103 (0·393)	38	41 [5–114]
4	0·0573 (0·713)	40	79 [10–219]

As shown in Table 3, the estimated effective size of the population based on the temporal variation approach (for each line over the nine generations), N_{eV}, was always higher than the estimate of the effective size based on the pedigree approach, N_{eI}. This indicated that the MHC locus was globally evolving at a lower rate than expected based on the pedigree data. The probability of obtaining in the simulated distribution an f_c, value equal to or greater than the observed f_c was high for the three selected lines (P value = 0·599, 0·663 and 0·393 for lines 1, 2 and 3, respectively) as well as for the control line (P value = 0·713), indicating that the four observed f_c values were not extreme, compared with the simulated f_c distribution.

5. Discussion

The estimate of the MHC effect on the three selected immune traits (ND3, PHA and CC) varied according to the method used. What lessons may be drawn from this study?

(i) *The interest in combining complementary analysis*

The interest in considering the MHC effect as random was twofold. First, it reduced the effect of rare genotypes, since in particular the initial frequency of the B^{34} haplotype was low, and second, it allowed consideration of the effect of the whole locus. On the contrary, assuming MHC as a fixed effect allowed us to have an approach focused on haplotypes instead of genotypes, and more particularly with the emphasis laid on each haplotype. Comparing the two

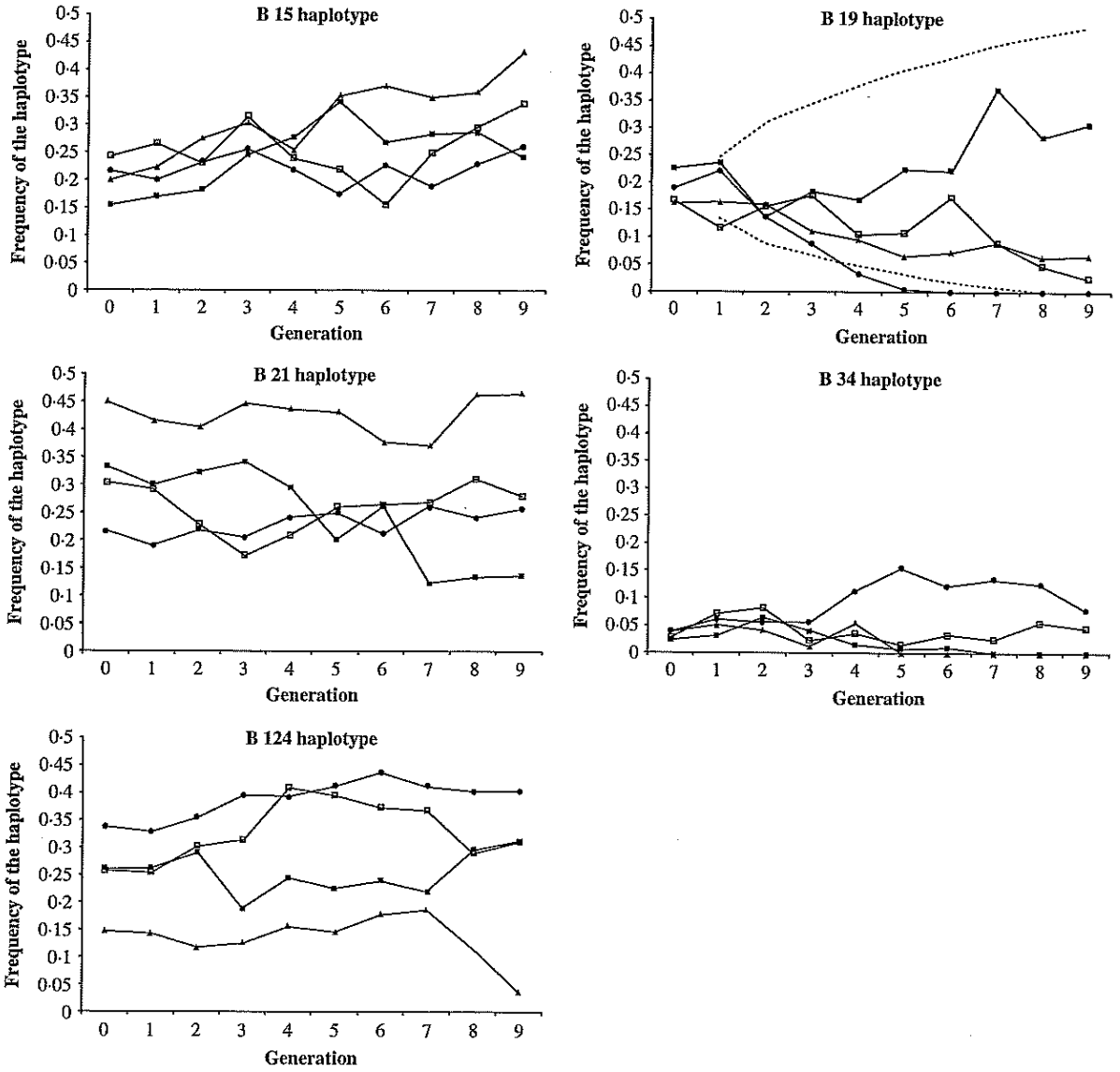


Fig. 1. Evolution of the observed frequencies of the haplotypes B^{15}, B^{19}, B^{21}, B^{34} and B^{124} within each selected line (unbroken lines: line 1, circles; line 2, squares; line 3, triangles) and within the control line (line 4, empty squares). Bounds of the 95% confidence interval are drawn (dotted lines) for the B^{19} haplotype within line 1.

hypotheses (fixed effect vs random effect) let us know whether the locus-scale variations detected when MHC is considered as a fixed effect may be large enough to be detected at the genome scale. Indeed, the negative effect of the B^{19} haplotype on trait ND3 was confirmed whereas the negative effect of the B^{124} haplotype on trait CC was not.

Likewise, the temporal method relies on the assumption that all variance in allele frequencies is due only to drift (Waples, 1989) and considers the locus as a whole. Effect of selection on the B^{19} haplotype within line 1 (trait ND3), revealed by the evolution of the haplotype frequency, was not strong enough to be detected using the temporal variation method. However, one limitation of the temporal method was shown by Pollack (1983): selection of constant intensity has a minor effect on f if t/N_e is small, which is the case in our study. In addition, in a multiallelic case, it might not detect even large variation in one specific allele frequency, provided the rest of the variation is alloted between the other different alleles. Therefore, the approach of the evolution of haplotype frequencies, haplotype by haplotype, is more efficient for detecting any impact of selection when the selection rules are maintained over generations and when a specific haplotype is selected.

Finally, evolution of the polymorphism at the MHC locus has shown that MHC had an effect on trait ND3 and the significant effect was confirmed by the variance components approach. This trait deals

with the antibody production against an antigen: Newcastle disease virus. Such responses against a variety of pathogens were already found to be correlated with MHC, as for instance the response against *Salmonella* bacteria (Guillot *et al.*, 1995) or against Marek's disease virus (Bacon, 1987). However, the "background genome also has a substantial effect" (Lamont, 1998) and most studies rely only on the observation of differences in allelic frequencies between lines to deduce an association between the trait and the MHC locus. In some cases, complementary analyses such as the analysis of the variance components would allow separation of the effect of the MHC gene from the non-MHC gene effects. In this study, only the evolution of the polymorphism would lead us to conclude the effect of the B^{34} haplotype on PHA and CC, since this haplotype disappeared in lines 2 and 3, and the complementary methods have not shown it. Concerning the B^{124} haplotype, complementary analyses have not given a clear picture of the effect of this haplotype on trait CC but have laid the emphasis on contradictory forces in the evolution of the B^{124} haplotype frequency. Indeed, the variance components approach has revealed a significative negative effect of the B^{124} haplotype on trait CC, which should lead to a decrease in the haplotype frequency, as happened within line 1 for the B^{19} haplotype; however, the significantly higher effect of homozygotes over heterozygotes on trait CC may slow down the decrease in frequency and explain why the effect of the B^{124} haplotype was not detected at the locus scale.

(ii) *Estimation of effective population size*

The effective size (N_e) of the population is a key parameter for discerning allele frequency changes due to drift versus those due to selection and hitch-hiking. N_e is usually estimated from the rate of inbreeding (N_{eI}) or from the variance in allele frequencies over time (N_{eV}). Both methods were used in this study. The values of these two parameters N_{eI} and N_{eV} were expected to be equal since the two methods are based on the analysis of a neutral gene. However, as pointed out by Crow & Kimura (1970), N_{eI} is usually smaller than N_{eV} when a small number of parents generate a large number of offspring because 'the inbreeding effective number is more naturally related to the number of parents, while the variance effective number is related to that of the offspring'. Here, the two realized effective sizes were found to have different values in the experimental lines analysed, N_{eV} being in one case twice N_{eI}, but the above explanation may not be sufficient to explain such a large difference. Additionally, confidence intervals of N_{eV} in the four lines are rather large since there are only five haplotypes at a single locus. The method of calculation of the bounds of the 95% interval, provided by Nei & Tajima (1981) and used by Waples (1989), is 'only asymptotically correct' for a large number of independent alleles.

It seems that allele frequency variations at the MHC locus are weaker than those of the whole genome (see pedigree analysis). Combining the two approaches allows us to detect whether another force is needed in addition to drift to explain the evolution of the haplotype frequencies.

(iii) *Contrasting results on the selection response and the MHC locus effect*

Response to selection was significant but variable within the three selected lines (Loywyck *et al.*, 2005): the antibody response (ND3) was the trait with the highest and the most regular increase in its mean under selection and the highest estimated heritability (0·35), whereas the increase in the means of traits PHA and CC was low and fluctuating and their estimated heritability lower (0·13 and 0·15, respectively). Then, assessing the effect of the MHC locus may be more difficult: if the response to selection of the trait is low, the evolution of the polymorphism at the candidate gene is reduced.

Assuming MHC as a fixed effect raised the question about the underlying model, which supposes additivity, i.e. heterozygote genotypes showing intermediate values between the two corresponding homozygote genotypes. Here, the significantly lower values for the heterozygote genotypes compared with their homozygote counterparts led us to reject the underlying additive model.

The MHC locus is known to be extremely polymorphic and its variation is thought to be maintained by balancing selection either through heterozygous advantage or negative frequency selection (Hugues, 1998; Bodmer, 1972). However, both hypotheses are controversial (Slade & McCallum, 1992) and Takahata & Nei (1990) concluded that frequency-dependent selection (specifically, rare allele frequency) and overdominance could not be distinguished mathematically. The results of this study do not support the hypothesis of negative frequency-dependent selection, which assumes that it is advantageous to carry rare alleles to which pathogens are not adapted (Bodmer, 1972): here, the rare haplotype B^{34} disappeared in lines 2 and 3 although this haplotype had the highest effect on the two traits PHA and CC, respectively. As in a recent study on birds that looked for overdominance (De Boer, 2004), no evidence in favour of the hypothesis of a heterozygous advantage has been reported in the present study, since the significant difference between the effect of heterozygote genotypes compared with the effect of homozygote genotypes was in favour of homozygosity.

6. Conclusion

This study highlights the interest in using various sources of information and methods in analysing a complex phenomenon such as testing whether a candidate gene is neutral or not. Different but complementary points of view have to be handled by either considering the locus as a whole or by analysing each allele separately, so that results may be contrasted or nuanced.

References

Bacon, L. D. (1987). Influence of the MHC on disease and productivity. *Poultry Science* **66**, 802–811.

Bodmer, W. F. (1972). Evolutionary significance of the HLA system. *Nature* **237**, 139–145.

Bumstead, N., Millard, B., Barrow, P. & Cook, J. K. A. (1991). Genetic basis of disease resistance in chickens. *Breeding for Disease Resistance in Farm Animals* (ed. J. B. Owen & R. F. E. Axford), pp. 10–23. Melksham, UK: CAB International.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.

De Boer, R. J., Borghans, J. A. M., van Boven, M., Kesmir, C. & Weissing, F. J. (2004). Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics* **55**, 725–731.

Goldringer, I. & Bataillon, T. (2004). On the distribution of temporal variations in allele frequency: consequences for the estimation of effective population size and the detection of loci undergoing selection. *Genetics* **168**, 563–568.

Goudet, J. (1995). FSTAT (version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity* **86**, 485–486.

Groeneveld, E. (1990). *PEST User's Manual*. Urbana, Illinois: University of Illinois.

Groeneveld, E. (1997). *VCE4 User's Guide and Reference Manual*. Neustadt, Germany.

Guillot, J. F., Beaumont, C., Bellatif, F., Mouline, C., Lantier, F., Colin, P. & Protais, J. (1995). Comparison of resistance of various poultry lines to infection by *Salmonella enteritidis*. *Veterinary Research* **26**, 81–86.

Hughes, A. L. & Yeager, M. (1998). Natural selection and the evolutionary of major histocompatibility complex loci. *Frontiers in Bioscience* **3**, 509–516.

Kaufman, J. F. & Lamont, S. J. (1996). The chicken major histocompatibility complex. In *The Major Histocompatibility Complex in Domestic Animal Species* (ed. L. B. Schook & S. J. Lamont), pp. 35–64. Boca Raton, Florida: CRC Press.

Lamont, S. J. (1998). Impact of genetics on disease resistance. *Poultry Science* **77**, 1111–1118.

Loywyck, V., Bijma, P., Pinard-van der Laan, M. H. & Verrier, E. (2005). A comparison of two methods for prediction response and rates of inbreeding in selected populations with the results obtained in two selection experiments. *Genetics Selection, and Evolution* **37**, 273–289.

Miller, M. M., Bacon, L. D., Hala, K., Hunt, H. D., Ewald, S. J., Kaufman, J., Zoorob, R. & Briles, W. E. (2004). Nomenclature for the chicken major histocompatibility (B and Y) complex. *Immunogenetics* **56**, 261–279.

Nei, M. & Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics* **98**, 625–640.

Pinard-van der Laan, M. H. (2002). Immune modulation: the genetic approach. *Veterinary Immunology and Immunopathology* **87**, 199–205.

Pollak, E. (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics* **104**, 531–548.

Slade, R. W. & McCallum, H. I. (1992). Overdominant vs. frequency-dependent selection at MHC loci. *Genetics* **132**, 861–862.

Sorensen, D., Guldbrandtsen, B. & Jensen, J. (2003). On the need for a control line in selection experiments: A likelihood analysis. *Genetics Selection and Evolution* **35**, 3–20.

Takahata, N. & Nei, M. (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**, 967–978.

Waples, R. S. (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379–391.

Weir, B. S. & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Wright, S. (1969). Evolution and the genetics of population. *The Theory of Gene Frequencies*, vol. 2. Chicago: University of Chicago Press.

Genet. Res., Camb. (2006), **87**, *p.* 223. © 2006 Cambridge University Press
doi:10.1017/S0016672306008275 Printed in the United Kingdom

Erratum

Loywyck, V., Pinard-van der Laan, M.-H., Goldringer, I. & Verrier, E. (2006). On the need for combining complementary analyses to assess the effect of a candidate gen and the evolution of its polymorphism: the example of the Major Histocompatibility Complex in chicken. *Genetical Research* **87**, 125–131.

The published version of Table 2 omitted the column headings, the correct version is given below.

Table 2. *Estimation of the effects of the MHC haplotypes on the three traits, using the method of contrasts and considering the MHC effect as fixed*

Haplotype	Trait ND3		Trait PHA		Trait CC	
15	0·175	*ab*	0·049	*a*	0·008	*bc*
19	0	*b*	0·008	*a*	0·006	*bc*
21	0·449	*a*	0	*a*	0·010	*ab*
34	0·441	*a*	0·050	*a*	0·033	*a*
124	0·356	*a*	0·047	*a*	0	*c*

Values are given considering the lowest effect as the reference, e.g. values of the effects of haplotypes on trait ND3 are given as compared to the value of the effect of B^{19} haplotype. Two haplotypes sharing the same letter (a, b or c) have effects that are not significantly different.

II. Evolution of genetic parameters over time.

The objective of this study was to investigate changes in additive genetic variance during the course of the experiment and examine the cope of the underlying model used for the estimation of the genetic parameters.

1. METHODS

Univariate analyses were carried out by the Restricted Maximum Likelihood (REML) method using VCE software (Groeneveld, 1997) on the three selected traits. For each trait the animal model was as follows:

$$Y_{ijk} = \mu + gener_i + sex_j + A_{ijk} + E_{ijk}$$

where Y is the performance, μ the overall mean, *gener* and *sex* are the environmental fixed effects of the generation and the sex of animals, respectively, A is the additive genetic value, assumed to be from polygenic origin, and E the random error.

Performance and pedigree data from all lines and all generations, up to the base generation, were available.

In a first step, the number of generations taken into account for the analysis was increased successively. In a second step, consecutive subsets of generation were taken: generation 0 to 2, generation 3 to 5 and the generation 6 to 9. All pedigree information available back to generation 0 was included for the analysis of each subset, estimations being underestimated when back pedigrees were omitted because parents without records would treated as if they were unrelated, i.e. less inbred than they really were (Meyer and Hill, 1991).

2. RESULTS AND DISCUSSION

For trait CC, estimates given in Tables 1 and 2 have to be nuanced because response to selection in the selected line for trait CC (Line 3) showed the largest variations from one generation to another (Loywyck *et al.*, 2005 – see chapter 2).

Table 2 shows evolution of estimates of genetic parameters over time for the 3 selected traits ND3, PHA and CC, increasing the number of generations to give data sets. For traits ND3 and PHA, estimates of the additive genetic variance (σ^2_A) fluctuated when only three generations

of data were considered but σ^2_A constantly decreased when more data from the later generations were added. Estimates of the residual variance (σ^2_E) followed the same trend as the additive genetic variance for trait ND3, whereas σ^2_E largely increased when adding data from the earlier generations for trait PHA. Then, estimated heritability (h^2) decreased over time for traits ND3 and PHA.

Table 3 shows evolution of estimates of genetic parameters 3 selected traits ND3, PHA and CC, based on subset of data. For traits ND3 and PHA, estimates of the additive genetic variance and of the residual variance decreased when considering latter generations. Decrease of σ^2_E being larger than σ^2_A for trait ND3 (24% and 31% for σ^2_A and σ^2_E between subset G0-G2 and subset G6-G9, respectively) estimate heritability increased; on the contrary, decrease of σ^2_E being smaller than σ^2_A for trait PHA (44% and 38% for σ^2_A and σ^2_E between subset G0-G2 and subset G6-G9, respectively) estimated heritability decreased. Environmental sensitivity seems to decrease in the latter generations.

The underlying model for the analysis of the three selected traits, using REML method, was the infinitesimal additive model, which assumes that trait are determined by infinity of unlinked additive genes with small effects and gene frequencies not affected by selection. Then, estimations of σ^2_A over time, when increasing the number of generations into account, should be identical since genetic variances are expected to be constant over time, except as a result of inbreeding and of selection that may cause linkage disequilibrium among unlinked loci; although, additive genetic variance decreased for both traits. Meyer and Hill (1991) have shown by simulations using data from section experiment for food intake in mice that effects of selection on the genetic variance were negligible when heritability was low (h^2=0.15). When heritability was higher (h^2=40) selection induced a decrease of genetic variance in the early generations but the effect was much smoother in the later generations. Thus, variation causes of the additive genetic variance for traits PHA and CC may be explained by increase of inbreeding within the lines whereas additive genetic variance for traits ND3 could be influenced by inbreeding but also affected by the selection strategy in the experiment.
Estimation of genetic parameters using data from all generations (G0-G9) was slightly different from estimation given by Loywyck *et al.* (2005) since previous estimations were handled considering only data from the line selected for the trait and from the control line.

Sorensen and Kennedy (1986) obtained unbiased estimates of σ^2_A when ignoring data from the first three generations in a selection experiment with 5 generations, which surprising since information contributing to selection decisions has been ignored. However, extending the simulations for up to 10 generations, van der Werf and de Boer (1990) showed that omitting data from selected ancestors caused bias in estimate of σ^2_A due to not accounting for all gametic disequilibrium. Linkage would be expected to lead to a progressive reduction in genetic variance over early generations: this would explain the decrease of σ^2_A when considering subset of data but not the large decline when considering the latter generations.

This study shows that selection for the immune response traits has reduced the genetic variance over and above the effects of inbreeding and linkage. The interest of the approach proposed by Meyer and Hill (1990) lies in the inferences that we are able to drawn on changes in variance over time that were not due simply to inbreeding or short-term effects of selection.

Table 2: Evolution of estimates of genetic parameters over time for the three traits, increasing the number of generations to give performances data sets (all pedigrees considered).

Data set	trait ND3			trait PHA			trait CC		
	σ^2_A	σ^2_E	h^2	σ^2_A	σ^2_E	h^2	σ^2_A	σ^2_E	h^2
G0-G1	1.087	2.028	0.349	0.085	0.265	0.236	0.002	0.010	0.167
G0-G2	1.398	2.329	0.382	0.091	0.444	0.179	0.002	0.011	0.154
G0-G3	1.284	2.325	0.357	0.074	0.479	0.134	0.002	0.012	0.143
G0-G4	1.436	2.395	0.374	0.065	0.554	0.105	0.001	0.011	0.083
G0-G5	1.300	2.267	0.362	0.070	0.523	0.118	0.002	0.010	0.167
G0-G6	1.187	2.241	0.344	0.073	0.496	0.127	0.002	0.010	0.167
G0-G7	1.169	2.097	0.357	0.073	0.471	0.131	0.003	0.009	0.250
G0-G8	1.133	2.095	0.349	0.063	0.451	0.119	0.003	0.009	0.250
G0-G9	1.075	2.065	0.339	0.056	0.427	0.112	0.003	0.009	0.250

Table 3: Evolution of estimates of genetic parameters for the three traits based on different subsets of data.

Data set	trait ND3			trait PHA			trait CC		
	σ^2_A	σ^2_E	h^2	σ^2_A	σ^2_E	h^2	σ^2_A	σ^2_E	h^2
G0-G2	1.398	2.329	0.382	0.091	0.444	0.179	0.002	0.009	0.182
G3-G5	1.364	2.105	0.385	0.082	0.564	0.126	0.002	0.009	0.182
G6-G9	1.058	1.608	0.391	0.051	0.277	0.145	0.004	0.008	0.333

CHAPTER 4: Analysis and comparison of the evolution of polymorphism of molecular markers

This chapter deals with analysis of genetic diversity based on molecular information from various markers, some markers being supposed neutral, the others supposed to be under selection. All markers were genotypes in the French experiment.

The goal of this chapter is to analyse and compare of the polymorphism evolution of different markers but also to investigate what are the suitable methods to detect any signature of selection left by QTLs. We aim at combining and comparing different methods to highlight selection pattern, QTL having already been detected in the Dutch experiment.

Genetics Selection Evolution, 2008, in press

Evolution of the polymorphism at molecular markers in QTL and non-QTL regions in selected chicken lines

Valérie LOYWYCK[1*], Bertrand BED'HOM[1],
Marie-Hélène PINARD-van der LAAN[1] , Frédérique PITEL[2],
Etienne VERRIER[1], Piter BIJMA[3]

[1] Génétique et Diversité Animales UMR1236, INRA / AgroParisTech,
78352 Jouy-en-Josas, France
[2] Génétique Cellulaire UR444, INRA, 31326 Castanet-Tolosan, France
[3]Animal Breeding and Genomics Group, Wageningen University, 6700 AH
Wageningen, The Netherlands

* Corresponding author. Tel: +33 1 34652947. Fax: +33 1 34652210.
e-mail: valerie.loywyck@agroparistech.fr

Short title: signature of selection in chicken

Abstract - We investigated the joint evolution of neutral and selected genomic regions in three chicken lines selected for immune response and in one control line. We compared the evolution of polymorphism of 21 supposedly neutral microsatellite markers versus 30 microsatellite markers located in seven Quantitative Trait Loci (QTL) regions. Divergence of lines was observed by factor analysis. Five supposedly neutral markers and 12 markers in the QTL regions showed *Fst* values greater than 0.15. However, the non significant difference (P>0.05) between matrices of genetic distances based on genotypes at supposedly neutral markers on the one hand, and at markers in QTL regions, on the other hand, showed that none of the markers in the QTL regions were influenced by selection. A supposedly neutral marker and a marker located in the QTL region on chromosome 14 showed temporal variations in allele frequencies that could not be explained by drift only. Finally, to confirm that markers located in QTL regions on chromosomes 1, 7 and 14 were under the influence of selection, simulations were performed using haplotype dropping along the existing pedigree. In the zone located on chromosome 14, the simulation results confirmed that selection had an effect on the evolution of polymorphism of markers within the zone.

Keywords: selection, QTL, hitchhiking, chicken, genetic diversity

1. INTRODUCTION

There is currently a large interest in characterising variation patterns in order to identify regions of the genome that are under selection. For that purpose, scans using microsatellites distributed over a genome [32, 35] or concentrated around candidate genes under artificial or natural selection [2, 28, 43] are commonly performed to investigate signatures of selection. These studies highlight and compare among natural populations, differences in patterns of heterozygosity or linkage disequilibrium, but they only give a picture of variability at a certain time, with predictions of the evolution of polymorphism estimated mainly through simulations. Well-known pedigree experimental selected lines can be used to explore the evolution of polymorphism over several generations, leading to the introduction of a time component that helps to distinguish the influence of selection from the influence of drift.

Here, we investigate the joint evolution of neutral and selected genomic regions, using observations on microsatellite markers in a number of selected chicken lines. For this purpose, we compared the evolution of marker allele frequencies observed in supposedly neutral versus selected regions of the genome. Selected regions were chosen based on Quantitative Trait Loci (QTL) detected in previous studies. An important aim was to

determine which methods are suitable for identifying signatures of selection, and to compare those methods using a real data set.

2. MATERIAL AND METHODS

2.1. Selection design

We used four experimental chicken lines bred since 1994 in the INRA experimental unit "*Unité Expérimentale de Génétique Factorielle Avicole*" (Nouzilly, France) and derived from an unselected base population of White Leghorn chickens [31] for which 42 founder animals of two lines (9 sires of a commercial line and 33 dams of an experimental line) were randomly mated (Generation G-2). The F2 population has become the base population, also named Generation 0 (G0). Animals from G0 were randomly chosen to create the four lines, thus the parents of one line cannot be parents of another line.

Three of these lines were selected for high values according to three different criteria of immune response: antibody response three weeks after vaccination against the Newcastle disease virus (line 1, trait ND3), cell-mediated immune response at nine weeks of age (line 2, trait PHA), and phagocytic activity at 12 weeks of age (line 3, trait CC). The three lines have undergone mass selection with a restriction on the contribution of the

different families (sizes of the different half-sib families were approximately balanced). The fourth line was the control line, in which the parents were chosen at random.

Within each line and at each generation (one generation per year), 15 males and 30 females out of about 100 candidates of each sex were chosen as parents for the next generation. Mating was at random, except that full- and half-sib mating was avoided. This selection programme was conducted for 11 discrete generations (G1 to G11). All animals of the four lines were measured for the three traits. The pedigree was completely known.

Estimated heritabilities were 0.33, 0.12 and 0.24 for the traits ND3, PHA and CC, respectively, using pedigree and phenotypic data up to generation 9 [22]. For other detailed results on this experiment, including genetic gains, various criteria of genetic variability and evolution of the polymorphism at a single candidate gene, namely the Major Histocompatibilty Complex (*MHC*) gene, see [21, 22].

2.2. Genotyping

In order to compare the evolution of polymorphism of supposedly neutral areas and selected areas, we decided to compare the evolution of microsatellites from the Aviandiv panel (European project on the analysis of

diversity in the chicken) and the evolution of microsatellites located within QTL regions, previously detected in independent studies on other lines.

2.2.1. Sampling of animals to be genotyped

Due to financial constraints, it was not possible to genotype animals in each generation. From G-2, 37 founders out of 42 were genotyped because blood samples from five founders were either missing or improper for DNA extraction. To reconstruct the five missing genotypes, and to determine the phase of haplotypes in QTL regions, 55 animals from generation G-1 were genotyped. Fifty animals of each line from G11 randomly chosen within half-sib families were genotyped.

2.2.2. Markers

The supposedly neutral markers are a set of di-nucleotide microsatellite markers used in a project on the biodiversity of chickens funded by the European Commission, namely known as the Aviandiv project [15]. These are distributed as uniformly as possible throughout the chicken genome. The position of the markers is given in Appendix 1 (published in electronic form only).

QTL-regions affecting the immune response were primo-detected in two other experimental lines bred on the experimental unit of the Animal breeding and Genomics Group at the Wageningen University and Research Center (Netherlands) [36, 37, 38]. The first population was an F2 originating from a cross of divergently selected lines for high and low antibody response to sheep red blood cells [42]. The second population was an F2 originating from a cross between two commercial lines [3]. Among the different regions detected, we chose six genome-wide significant QTL regions for different antibody titre traits. The presence of these QTL was not checked in our experimental lines due to financial constraints, which limited the number of genotyped animals. The *MHC* region (chromosome 16 – zone 7) was added to the analysis, since the *MHC* gene is a good candidate gene for immune response [22].

The distance between markers was defined according to estimations of allele frequency changes of markers under selection in mouse lines [18] and estimation of the extent of linkage disequilibrium in domestic sheep [23], since such estimations have not been conducted in chicken. The position of the markers is given in Table I. Genetic distances of existing markers were those defined by the consensus map [12] and genetic distances of the new markers were estimated from the consensus map and their position on the chicken genome sequence.

The genetic position of the three markers within zone 7 (*MHC* region) was found to be the same (~ 0 cM) on the consensus map: in order to run simulations, positions were arbitrarily set to 0.00, 0.05 and 0.10 cM in the strict case of this study.

[Tab. I]

Fluorescently labelled microsatellite markers were analysed on an ABI 3100 DNA sequencer (Applied Biosystems, Foster City, CA, USA) and genotypes were determined using GeneScan Analysis 3.7 and Genotyper Analysis 3.7 software (Applied Biosystems, Foster City, CA, USA). The GEMMA database was used to manage the informativity tests [16]. A recent analysis (Bed'Hom – *unpublished results*) of the markers located in the *MHC* region (zone 7) revealed the presence of a null allele for *MCW370*. The null allele was named AAA and genotypes were rebuilt according to specific associations of marker alleles within the zone. Appendices 2 and 3 summarise the observed allele frequencies in G-2 and G11 (Appendices 2 and 3 are available in electronic form only).

2.3. Measures of line divergence

2.3.1. Factor analysis

In order to get an overview of the distinction among generations and among lines, we performed a multiple-dimension principal component analysis (PCA) on all individuals, from generations G-2, G-1 and G11. First, PCA was based on genotypes at all markers. Second, in order to assess the influence of the different types of markers, PCA was based, on genotypes at the supposedly neutral markers, on the one hand and on genotypes at markers in QTL regions, on the other hand.

2.3.2. Genetic variability criteria

In order to quantify genetic differences between the lines, we calculated standard descriptors of the genetic variability for each locus in G-2 and in G11 within each line: observed heterozygosity H_0 and unbiased expected genetic diversity H_{exp} [29]. Departures from Hardy-Weinberg equilibrium were estimated by calculating Wright's *Fis* and *Fst* according to Weir and Cokerham [45]. The null hypothesis (F_{is}=0) was tested by bootstrapping over alleles within samples. Population differentiation was tested by permuting genotypes among samples, assuming absence of Hardy-Weinberg equilibrium within samples.

Pairwise linkage disequilibrium was estimated by testing the significance of association between genotypes at pairs of loci within QTL regions and across supposedly neutral loci; this analysis was performed in

G-2 and G11 within each line. P-values were obtained by randomisation of the genotypes at each pair of loci. In order to take into account the fact that multiple loci were examined, a Bonferroni correction was applied within each line. Calculations dealing with heterozygosity and linkage disequilibrium were performed using the F-STAT programme [11].

In order to quantify the genetic divergence over time of our lines deriving from the founder population, we estimated the genetic distances. We assumed that mutations at the microsatellite markers could be neglected. It has been reported that divergence occurred on a short-term period and inbreeding increased within each line [21]. Thus, the Reynolds distance [34] is preferred because under the assumption of pure genetic drift, it is the least biased genetic distance for closely related breeds and exhibits the smallest standard error [20]. Since our different markers are polymorphic loci with balanced or unbalanced allele frequencies in the founder population, we used weighted estimates of Reynolds distance, $\hat{D}_R^*$ [20]. The standard error of the weighted Reynolds distance, $\sigma(\hat{D}_R^*)$, is equal to:

$$\sigma(\hat{D}_R^*) = \sqrt{\frac{2}{\sum_{j=1}^{L}(k_{0,j}-1)}(\overline{F} + (1/n_0 + 1/n_t))} \qquad (1)$$

where $k_{0,j}$ is the number of alleles at the jth locus in the founder generation, n_0 and n_t are respectively the number of alleles in the founder generation and in generation G11 and $\overline{F}$ the average inbreeding coefficient [20]. Here,

weighted estimates of Reynolds distance and standard errors were computed between the G-2 population and lines in G11, and across lines in G11, using the POPULATIONS programme [19]. In order to assess the influence of the different types of markers, genetic distances were estimated using genotypes at supposedly neutral markers, on the one hand and genotypes at all markers in QTL regions, on the other hand.

2.4. Evolution of marker polymorphism within lines

2.4.1. Temporal changes in allele frequencies

In order to detect markers for which the evolution of polymorphism departs from evolution under pure drift, we estimated temporal changes in allele frequencies for each locus.

An estimate of the standardised temporal variance in allele frequency, f [47], was computed for each locus within each line over the 13 generations; the fc estimator of f, proposed by Nei and Tajima [30] was used:

$$\hat{f}_c = \frac{1}{k}\sum_{i=1}^{k}\frac{\left[x_{0,i} - x_{t,i}\right]^2}{\frac{x_{0,i} + x_{t,i}}{2} - x_{0,i} \cdot x_{t,i}} \qquad (2)$$

where k is the number of segregating alleles, $x_{0,i}$ is the frequency of allele i in G-2 and $x_{t,i}$ the frequency of this allele in G11. The observed value of fc

was compared to the distribution of *fc* obtained from simulations of populations under drift, with the same initial allele frequencies and the same inbreeding effective size [10]. P-values were computed for each locus. Because multiple loci were examined, expected false discovery rates, also known as *Q-values*, were calculated within each line using the QVALUE package [39]. The false discovery rate is the expected proportion of false positives among the tests found significant. A false positive is the term used to describe rejection of the null hypothesis (*i.e.*, calling the test significant) when it is really true. We fixed the false discovery rate at a pre-determined level of $\alpha = 5\%$ beforehand, in order to guarantee that the number of false positives would represent 5% or less of the number of significant tests.

The estimate of the variance effective size (Ne_V) of each selected line was directly deduced from the value of $\bar{f}_c$, using the equation of Waples [44]:

$$\hat{Ne}_V = \frac{t}{2[\bar{f}_c - 1/(2S_0) - 1/(2S_t)]} \quad (3)$$

where S_0 and S_t are respectively the sample sizes in the founder generation (G-2) and in generation G11, t is the number of generations and $\bar{f}_c$ is the mean of *fc* across the different loci, weighted by the number of alleles [40]. This value was compared to the value of effective size calculated from the pedigree, $Ne_I = 1/2\Delta F$

2.4.2. Simulations

In order to detect markers undergoing selection, we simulated the evolution of polymorphism of the different markers along the existing pedigree. Simulations (1000 iterations) using haplotype-dropping along the pedigree were performed. From the simulation iterations, a 95% confidence interval (CI) was drawn for the allele frequencies of each marker.

Initialisation: A haplotype consisted in the different markers located within a defined zone. Haplotypes in the selected zones and genotypes at the supposedly neutral markers were known for the 43 individuals of generation G-2. We drew different assumptions about QTL location in one of the selected zones and in that case, the favourable allele Q in generation G-2 was either defined as linked to a marker allele within the zone, or settled according to a given initial frequency.

Transmission: The approximate mutation rate in our dataset was calculated based on the number of new alleles in G11 (and confirmed with simulations), which yielded a mutation rate of 10^{-7}. Therefore, a stepwise mutation model was used with a 10^{-7} mutation rate. Recombination within the haplotype followed the Haldane model. Haplotypes and genotypes were

dropped along the existing pedigree conditional on the observed phenotypes.

First, we tested the assumption of pure drift: transmission of haplotypes and genotypes followed Mendelian transmission rules. Second, we assumed the presence of QTL related to one of the three traits in one of the QTL regions and tested the assumption of both selection and drift: transmission of genotypes and haplotypes in zones without QTL followed Mendelian transmission rules whereas transmission of the haplotype in the zone with the QTL was conditional to the transmission of the QTL. Transmission of the QTL was conditional on the phenotype of the offspring and on the QTL genotypes of the parents. In that case, we used the Bayes theorem:.

$$p(G_i/z) = \frac{p(G_i) \cdot p(z/G_i)}{\sum_j p(G_j) \cdot p(z/G_j)} \qquad (4)$$

where $p(G_i/z)$ is the probability that offspring inherit QTL genotype G_i given its phenotypic value z. The so-called prior probabilities of the three QTL genotypes, $p(G_1)=p(QQ)$, $p(G_2)=p(Qq)$ and $p(G_3)=p(qq)$ were calculated according to the genotypes of the parents. Probabilities of the phenotype given the QTL genotype, also called penetrance, were given by $p(z/G_i) = \varphi(z, \mu_i, \sigma^2)$, where μ_i is the phenotypic mean for the genotype i at the QTL, and σ is the phenotypic standard deviation (estimated in the base population, i.e., in generation G0). The distribution of the phenotype was

assumed to follow a normal distribution. We set the QTL values for the trait to $+\alpha$, $(k^*\alpha)$ and $-\alpha$ for genotypes QQ, Qq and qq respectively, k being the degree of dominance, using the same scale as Falconer and Mackay [8].

3. RESULTS

3.1. Line divergence

3.1.1. Factor analysis

A two-dimension analysis of all individuals based on genotypes of all markers discriminated individuals from G11 (Figure 1a). The three selected lines were distinct and well distributed, although the control line overlapped with individuals from generations G-2 and G-1 in the middle of Figure 1a. The first two principal components explained in total 10% of the variance.

We obtained the same picture when using only the genotypes of markers in the QTL zones but not when using the genotypes of supposedly neutral markers (Figure 1b and 1c): for supposedly neutral markers, individuals from G11 gathered at the centre and individuals from line 3 and from the control line overlapped.

[Fig.1]

A three-dimension analysis of all individuals based on genotypes of all markers showed that individuals from generations G-2 and G-1 were in a different plane than individuals from G11 (results not shown): the third axis seems to represent time divergence between generations G-2 and G11.

3.1.2. Genetic variability and genetic distances

Fis values of six markers (one supposed to be neutral and five in QTL zones) in G-2 were significantly different from zero, all markers showing an excess of heterozygosity. Excess of heterozygosity at the markers was observed for female founders originating from an experimental line with very few reproducers: in that case, allele frequencies are different for sires and for dams [33]; the more heterozygosity is in excess, the smaller is the number of reproducers. This excess was not observed anymore in G11. However, two markers showed a significant heterozygote deficiency: *SEQALL427* (zone 3) in line 1 and *ADL327* (zone 5) in lines 1 and 2. The supposedly neutral marker *ADL278* showed a significantly negative *Fis* value in G11 in line 2, whereas this marker did not show any departure from Hardy-Weinberg equilibrium in G-2. The results of deviations from Hardy-Weinberg equilibrium as estimated by *Fis* vales are presented in Table II.

[Tab. II]

Fst values ranged from 0.035 to 0.409. According to the Wright criterion, the important diversification (*Fst*>0.15) among lines in G11 was due to five supposedly neutral markers and 12 markers located in QTL zones. Estimated *Fst* values (and standard deviation) of those markers are presented in Table III.

[Tab. III]

No linkage disequilibrium between pairs of neutral loci was found, neither in generation G-2 nor in G11. On the contrary, significant linkage disequilibrium occurred between pairs of loci within each QTL zone. Linkage disequilibrium was also tested between the selected markers across the zones: linkage was only observed between markers within a given zone (detailed results not shown). Furthermore, in our simulations, these results will allow us to consider the supposedly neutral markers as independent whereas markers located in a QTL zone will consist in a haplotype.

Table IV gives the matrices of weighted Reynolds distances between the G-2 population and the four lines in G11, estimated either with genotypes of the supposedly neutral markers (upper matrix), or with genotypes of markers located in all QTL zones (lower matrix). Genetic distances between G-2 and any of the four lines in G11 tend to be larger using genotypes of markers located in all QTL zones than using genotypes of supposedly neutral markers. However, the Mantel test did not show a

significant difference between the two matrices, whether individuals from the control line were taken into account or not ($P>0.05$).

[Tab. IV]

3.2. Evolution of marker polymorphism within lines

3.2.1. Temporal variations in allele frequencies

Two markers show variations in allele frequencies that could not be explained only by drift: *fc* of the supposedly neutral marker *ADL278* was 0.559 (*Q-value* = 0.01) in line 3 and 0.324 (*Q-value* = 0.00) in line 4; *fc* of *SEQALL454* in zone 2 was 0.485 (*Q-value* = 0.00) in line 4. For loci for which variations could be explained by drift, the average *fc* value was 0.135 (±0.101).

3.2.2. Simulations

The 95% confidence intervals (CI) were very large under the assumption of pure drift. The observed allele frequencies of six markers (in zones 1, 2 and 3) fell outside the 95% CI. The observed allele frequencies and 95% CI of those markers are given in Table V.

[Tab. V]

There is no multiple testing in the results of simulations, but considering the total number of alleles per zone, we may approximate the expected number of false positives. The expected number of false positives is four for zones 1 and 2 and three for zone 3. The number of observed allele frequencies that fall outside the 95% CI is larger than the expected false positives for zone 2. Consequently, and according to previous results about genetic variability, we shall focus on zone 2 in greater detail.

QTL in zone 2 was primo-detected for antibody titre to KLH and *M. butyricum*, which are complex antigens. Such complex antigens bind to Th1 or Th2 cytokines and lead to a combination of cellular and humoral mediated pathways [9, 17]. Trait PHA corresponds to the cell-mediated immune response. To understand the evolution of markers located in this zone, different assumptions were drawn about the presence of a QTL affecting trait PHA (*i.e.*, the selected trait in line 2). First, we compared the observed allele frequencies in G11 in the four lines. Second, we confronted the genotypes of individuals at each marker with the lowest and the highest PHA phenotypes. This gave us indications on any particular association between the marker alleles and the QTL alleles. Then, we tested different localisations of the QTL within zone 2, different degrees of dominance between the QTL alleles and different effects of the QTL on trait PHA. However, the observed allele frequencies of *SEQALL455* never fitted the

95% CI drawn under the different assumptions about a bi-allelic QTL simulated within zone 2.

Further investigation of genotyping results led us to question the real polymorphism of two markers, namely *SEQALL453* in zone 2 and *ADL327* in zone 5: for both of them, a pseudo-null allele seems to exist (with a size of 209 bases for *SEQALL453* and 107 bases for *ADL327*) and was not detectable according to the other allele in the genotype. These assumptions may offset the effects of selection on these markers.

3.3. Effective population size

Table VI shows the estimations of the effective size for each line, based on the rate of inbreeding using pedigree information (Ne_I) or based on variations of allele frequencies (Ne_V) either from supposedly neutral markers or from markers in all QTL zones.

[Tab. VI]

The values obtained via the temporal variation approach (Ne_V) were always higher than the values derived from the rate of inbreeding (Ne_I). Moreover, Ne_V values estimated either from supposedly neutral markers or from markers in QTL zones were significantly different everywhere except in the control line. The value from neutral markers was significantly lower than the one from markers in QTL zones in lines 1 and 2, and the opposite

was observed in line 3. It has to be noted that, in the three selected lines, estimations of the effective size based on temporal allele frequencies at the *MHC* locus [22] were equivalent to Ne_F using genotype information from markers in all QTL zones, but estimated values were larger in the control line *i.e.* 76 for the control line and 51, 65 and 41 for lines 1, 2 and 3, respectively.

4. DISCUSSION

4.1. Combining different methods for the detection of signature of selection

Factor analysis gives a good overview of the divergence of lines and constitutes an interesting starting point in detecting signatures of selection. The non-significant difference between matrices of genetic distances, according to the type of markers considered, let us suppose that not all markers in the QTL zones are influenced by selection. The evolution of polymorphism of loci over time (fc) and fixation indices allowed us to focus on a smaller set of markers that may be influenced by selection.

Finally, to confirm which marker was actually under the influence of selection, simulations were performed since they could take the selection scheme into account (the pedigree was completely known).

4.2. Improving the detection of signature of selection

The extent of selective sweep and the distortion in allele frequency spectrum depend on the strength of selection and time since selection occurred *e.g.* [1, 4] but also on original marker variability and marker density. In our experiment, the strength of selection was attenuated since we tried to balance the representation of the half-sib families. The low marker density in our dataset was partly due to the limited number of microsatellites known in the chicken genome and the limited number of polymorphic markers in our experimental lines. In chicken, dropping simulations along the pedigree would probably be more efficient using high-density genotypes. For instance, simulation results on bovine chromosomes [13] suggest that the signature of selection can be detected up to 1 Mb (assuming 1Mb ~ 1 cM) from a QTL. However, this effect may extend further since Pollinger *et al.* [32] showed a 40 Mb-selective sweep around a gene with a large phenotypic effect in dog (*i.e.* the *TYRP1* gene known to be responsible for black coat colour).

To improve detection of the signature of selection in our experimental lines still using our microsatellite markers, an earlier generation should be genotyped. Indeed, the number of crossing-overs increases with time and any particular association between a marker and a

potential QTL could be broken along the successive generations. This association could probably still be detected in earlier generations. This approach was confirmed by Wiener *et al.* [46] when comparing the effect of selection on *GDF-8* (myostatin gene associated with double muscling) in double muscled breeds, using microsatellite loci at various distances from *GDF-8*. Their study showed that selection on *GDF-8* had left a stronger mark in the breed in which the double-muscling mutation had been present for the shortest time.

4.3. Difficulties in detecting signature of selection on immune response traits

The results dealing with zone 2 (located on chromosome 14) agreed that selection had an effect on the evolution of polymorphism of markers within the zone. However, modelling selective sweep was not easy and the underlying model seems to be complex. A QTL may be involved in the evolution of polymorphism within this zone but not only, since the observed allele frequencies never exactly fitted the simulated confidence intervals. A polygenic background could be added or the presence of several QTL with low effects could be assumed with epistatic interactions within a zone, for instance. Crossbreeding (F1, F2 and backcrosses) created from generation G11 have been analysed for the three immune traits and the analysis showed

a significant recombination loss for ND3, which highlights the important epistatic interactions for this trait [26]. Pleiotropic effects of QTL on the three traits could also be considered, since the pairwise genetic correlations were shown to be non-significant [25, 31] but were still not null and the three traits represent different aspects of the complex mechanism of immune response.

Recent improvements in chicken genome mapping [27, 41] have shown a certain number of discordances that led us to question the genetic position but also the order of microsatellites located within zone 2. Such discordances do not disturb findings from statistical analyses but could disturb results from simulations.

QTL were primo-detected for primary antibody response to specific antigens such as Sheep Red Blood Cells (SRBC), *Mycobacterium butyricum* and Keyhole Limpet Hemocyanin (KLH), and for Lipopolysaccharide (LPS) natural antibodies. However, as in mammals, immune responses in avian species are specialised in the elimination of antigens: responses to antigens are Th1- or Th2-mediated [7]. Th1 responses require the interference of type 1 T helper cells that directs immune response toward a cell-mediated response (cellular pathway). Th2 responses require type 2 T helper cells that favour the development of humoral response (humoral pathway). KLH and SRBC antigens represent Th2-responses whereas *M. butyricum* represents Th1-response.

In our experimental dataset, line 1 was selected for antigens against ND3 (New Castle Disease virus), inducing a Th1-response [6] whereas traits selected in lines 2 and 3 deal with innate immune response. Markers from zone 2, primo-detected for antigens to KLH and *M. butyricum* and falling outside the 95% CI under assumption of pure drift in line 2, show that responses are rarely exclusively Th1 or Th2 mediated and even if immune responses to antigens follow the same pathway, there is additional complexity in the control of different antigens. The detected QTL were linked to immune response to specific antigens and could not match with our selected traits. This was confirmed by a recent experiment where antibody response to KLH, *M. butyricum* and LPS was tested in our experimental lines in generation G12 [24]: no difference was observed among lines for KLH and LPS antibodies, but line 1 selected for ND3 showed a significantly higher specific response to *M. butyricum*. Finally, this led us to retain the hypothesis that QTL may have not segregated in our experimental lines.

4.4. Effective population size

The effective size estimated from the rate of inbreeding (Ne_I) was slightly smaller than the effective size estimated from the variance of allele frequencies over time (Ne_V) of supposedly neutral markers. This agrees with

Crow and Kimura [5] who pointed out that Ne_I is usually smaller than Ne_V when a small number of parents generate a large number of offspring, with both estimations assuming neutrality of the markers. However, a surprising result was that estimation of effective size based on allele frequency variation from G-2 to G11 of markers located in QTL zones was larger than estimation from supposedly neutral markers for lines 1 and 2. This may be explained by selection acting like a backmoving force that draws allele frequencies in the same direction, whatever the selected line; in that case, fluctuations for allele frequencies are lower than for neutral loci *e.g.*, [14]. Another explanation may be that samples are taken from extreme generations and a calculation based on temporal variation in allele frequencies does not take into account fluctuations that occur over generations: samples from intermediate generations would have given more information.

It seems that allele frequency variations at the supposedly selected markers are weaker than those of the whole genome, as for the *MHC* locus, which is involved in different stages of the immune response [22]. Could this indicate that variations of markers that influence ND3 or PHA traits are maintained by balancing selection, like variations at the *MHC* locus, and that detection of signatures of selection when it deals with immunity traits is rather difficult? In addition, since experimental animals are vaccinated against other diseases, do these vaccinations have an impact on our trait

measures? This may explain why the observed allele frequencies of *SEQALL454* in zone 2 fall out of the confidence intervals even in the control line.

ACKNOWLEDGEMENTS

This study was supported jointly by the French Genetic Resources Bureau (BRG) and the scientific committee of AgroParisTech.

REFERENCES

[1] Barton N.H., Genetic hitchhiking. Philos. Trans. R. Soc. Lond. B Biol.Sci. 355 (2000) 1553-1562.

[2] Bauer DuMont V., Aquadro C.F., Multiple signatures of positive selection downstream of Notch on the X chromosome in *Drosophila melanogaster*, Genetics 171 (2005) 639-653.

[3] Buitenhuis A.J., Rodenburg T.B., van Hierden Y., Siwek M., Cornelissen S.J.B., Nieuwland M.G.B., Crooijmans R.P.M.A., Groenen M.A.M., Koene P., Korte S.M., Bovenhuis H., van der Poel J.J., Mapping quantitative trait loci affecting feather pecking behaviour and stress response in laying hens, Poult. Sc. 82 (2003) 1215-1222.

[4] Charlesworth B., Nordborg M., Charlesworth D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70 (1997) 155-174.

[5] Crow J.F., Kimura M., An introduction to population genetics theory, Harper & Row New York (NY), USA, 1970.

[6] Degen W.G.J, van Daal N., Rothwell L., Kaiser P., Schijns V.E.J.C., Th1/Th2 polarization by viral and helminth infection in birds, Vet. Microbiol. 105 (2005) 163-167.

[7] Erf G.F., Cell-mediated immunity in poultry, Poult. Sc. 83 (2004) 580-590.

[8] Falconer D.S., Mackay T.F.C., Introduction to quantitative genetics, Longman Scientific & Technical, Essex, UK, 1996.

[9] Gause W.C., Urban J.F. Jr, Stadecker M.J., The immune response to parasitic helminths: insights from murine models, Trends Immunol. 24 (2003) 269-277.

[10] Goldringer I., Bataillon T., On the distribution of temporal variation in allele frequency: consequences for the estimation of effective population size and the detection of loci undergoing selection, Genetics 168 (2004) 563-568.

[11] Goudet J. (2001), FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3), http://www.unil.ch/izea/sofwares/fstat.html

[12] Groenen. M.A.M., Cheng H.H., Bumstead N., Benkel B.F., Briles W.E., Burke T., Burt D.W., Crittenden L.B., Dodgson J., Hillel J., Lamont S., Ponce de Leon A., Soller M.,Takahashi H., Vignal A., A consensus linkage map of the chicken genome, Genome Res. 10 (2000) 137-147.

[13] Hayes B.J., Lien S., Nilsen H., Olsen H.G., Berg P., Maceachern S., Potter S., Meuwissen T.H.E., The origin of selection signatures on bovine chromosome 6, Anim. Genet. 39 (2008) 105-111.

[14] Hill W.G., Fixation probabilities of mutant genes with artificial selection, Genet. Sel. Evol. 17 (1985) 351-358.

[15] Hillel J., Groenen M.A.M., Tixier-Boichard M., Korol A.B., David L., Kirzhner V., Burke T., Barre-Dirie A., Crooijmans R.P.M.A., Elo K., Feldman M.W., Freidlin P.J., Mäki-Tanila A., Oosrtwijn M., Thompson P., Vignal A., Wimmers K., Weigend S., Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools, Genet. Sel. Evol. 35 (2003) 533-557.

[16] Iannucelli E., Woloszyn N., Arhainx J., Gellin J., Milan D., GEMM A: A database to automate microsatellite genotyping, in: Proceedings of the XXVth International Conference on Animal Genetics, 21-25 July 1996, Tours, France, Anim. Genet. 27, Suppl. 2-55.

[17] Jankovic D., Liu Z., Gause W.C., Th1- and Th2-cell commitment during infectious disease: asymmetry in divergent pathways, Trends Immunol. 22 (2001) 481-482.

[18] Keightley P.D., Bulfield G., Detection of quantitative trait loci from frequency changes of marker alleles under selection, Genet. Res. 62 (1993) 195-203.

[19] Langella O. Populations 1.2.28, logiciel de génétique des populations (2002). Laboratoire Populations, Génétique et Evolution, CNRS UPR9304, Gif-sur-Yvette, France.

[20] Laval G., San Cristobal M., Chevalet C., Measuring genetic distances between breeds: use of some distances in various short term evolution models, Genet. Sel. Evol. 34 (2002) 481-507.

[21] Loywyck V., Bijma P., Pinard-van der Laan M.-H., van Arendonk J., Verrier E., A comparison of two methods for prediction of response and rates of inbreeding in selected populations with the results obtained in two selection experiments, Genet. Sel. Evol. 37 (2005) 273-289.

[22] Loywyck V., Pinard-van der Laan M-H., Goldringer I., Verrier E., On the need of combining complementary analyses to assess the effect of a candidate gene and the evolution of its polymorphism: the example of the Major Histocompatibility Complex in chicken, Genet. Res. 87 (2006) 125-131. Erratum in: Genet. Res. 87 (2006) 223.

[23] Mc Rae A.F., McEwan J.C., Dodds K.G., Wilson T., Crawford A.M., Slate J., Linkage disequilibrium in domestic sheep, Genetics 160 (2002) 1113-1122.

[24] Minozzi G., Parmentier H.K., Nieuwland M.G.B., Bed'Hom B., Minvielle F., Gourichon D., Pinard-van der Laan M.-H., Antibody responses to keyhole limpet hemocyanin, lipopolysaccharide and Newcastle disease virus in F2 and backcrosses of White Leghorn lines selected for two different immune response traits, Poult. Sc. 86 (2007) 1316-1322.

[25] Minozzi G., Parmentier H.K., Mignon-Gastreau S., Nieuwland M.G.B., Bed'Hom B., Gourichon D., Minvielle F., Pinard-van der Laan M.-

H., Correlated effects of selection for immunity in White Leghorn chicken lines on natural antibodies and specific antibody responses to KLH and M. butyricum, BMC Genet. 9 (2008) 5.

[26] Minozzi G., Bidanel J.-P., Minvielle F., Bed'Hom B., Gourichon D., Baumard Y., Pinard-van der Laan M.-H., Crossbreeding parameters of general immune response traits in White Leghorn chickens, Livestock Sciences (2008) *in press*.

[27] Morisson M., Leroux S., Jiguet-Jiglaire C., Assaf S., Pitel F., Lagarrigue S., Bardes S., Feve K., Faraut T., Milan D., Vignal A., A gene-based radiation hybrid map of chicken chromosome 14: comparison to human and alignment to the assembled chicken sequence, Genet. Sel. Evol. 37 (2005) 229-251.

[28] Nair S., Williams J.T., Brockman A., Paiphun L., Mayxay M., Newton P. N., Guthman J-P., Smithuis F. M., Hien T. T., White N. J., Nosten F., Anderson T.J.C., A selective sweep driven by pyrimethamine treatment in southeast malaria parasites, Mol. Biol. Evol. 20 (2003) 1526-1536.

[29] Nei M., Molecular Evolutionary Genetics., Columbia University Press, New York, USA, 1987.

[30] Nei M., Tajima F., Genetic drift and estimation of effective population size, Genetics 98 (1981) 625-640.

[31] Pinard-van der Laan M.-H., Immune modulation: the genetic approach, Vet. Immunol. Immunopathol. 87 (2002) 199-205.

[32] Pollinger J.P., Bustanante C.D., Fledel-Alon A., Schmutz S., Gray M. M., Wayne R.K., Selective sweep mapping of genes with large phenotypic effects, Genome Res. 15 (2006) 1809-1819.

[33] Pudovkin A.I., Zaykin D.V., Hedgecock D., On the potential for estimating the effective number of breeders from heterozygote-excess in progeny, Genetics 144 (1996) 383-387.

[34] Reynolds J., Weir B.S., Cockerham C.C., Estimation of the coancestry coefficient basis for short-term genetic distance, Genetics 105 (1983) 767-779.

[35] Schlötterer C., A microsatellite-based multilocus screen for the identification of local selective sweeps, Genetics 160 (2002) 753-763.

[36] Siwek M., Cornelissen S.J.B., Nieuwland M.G.B., Buitenhuis A.J., Crooijmans R.P.M.A., Groenen M.A.M., de Vries-Reilingh G., Parmentier H.K., van der Poel J.J., Detection of QTL for immune response to Sheep Red Blood Cells in laying hens, Anim. Genet. 34 (2003) 422-428.

[37] Siwek M., Buitenhuis A.J., Cornelissen S.J.B., Nieuwland M.G.B., Bovenhuis H., Crooijmans R.P.M.A., Groenen M.A.M., de Vries-Reilingh G., Parmentier H.K., van der Poel J.J., Detection of different QTL for antibody response to keyhole lympet hemocyanin and Mycobacterium

1 butyricum in two unrelated populations of laying hens, Poult. Sc. 82 (2003)
2 1845-1852. Erratum in: Poultry Science 85 (2006) 1118-1119.
3 [38] Siwek M., Buitenhuis A.J., Cornelissen S.J.B., Nieuwland M.G.B.,
4 Knol E.F., Bovenhuis H., Crooijmans R.P.M.A., Groenen M.A.M., de
5 Vries-Reilingh G., Parmentier H.K., van der Poel J.J., Detection of QTL for
6 innate: non specific antibody levels binding LPS and LTA in two
7 independent populations of laying hens, Dev. and Comp. Immunol. 30
8 (2006) 659-666.
9 [39] Storey J.D., A direct approach to false discovery rates, J. Royal Stat.
10 Soc., Series B 64 (2002) 479-498.
11 [40] Tajima F., Nei M., Note on genetic drift and estimation of effective
12 population size, Genetics 106 (1984) 569-574.
13 [41] Takahashi H., Tsudzuki M., Sasaki O., Niikura J., Inoue-Murayama
14 M., Minezawa M., A chicken linkage map based on microsatellite markers
15 genotyped on a Japanese Large Game and White Leghorn cross, Anim.
16 Genet. 36 (2005) 463-467.
17 [42] van der Zijp A.J., Nieuwland M.G.B., Immunological
18 characterization of lines selected for high and low antibody production, 7th
19 European Poultry Conference, 1986, Paris, France, Vol.1, pp 211-215.
20 [43] Vigouroux Y., McMullen M., Hittinger C.T., Houchins K., Schulz
21 L., Kresovich S., Matsuoka Y., Doebley J., Identifying genes of agronomic
22 importance in maize by screening microsatellites for evidence of selection

during domestication, in: Proceeding of the National Academy of Sciences, 2002, Vol. 99, pp 9650-9655.

[44] Waples R.S., A generalized approach for estimating effective population size from temporal changes in allele frequency, Genetics 121 (1989) 379-391.

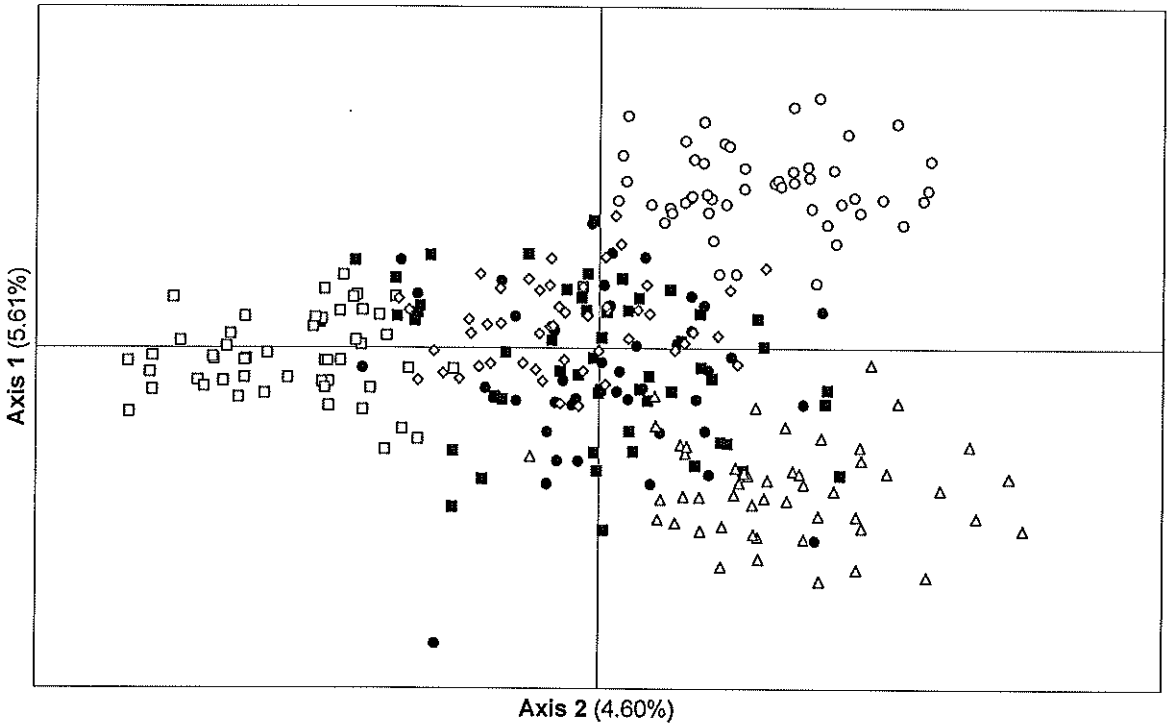
[45] Weir B.S., Cockerham C.C., Estimating F-statistics for the analysis of population structure, Evolution 38 (1984) 1358-1370.

[46] Wiener P., Burton D., Ajmone-Marsan P., Dunner S., Mommens G., Nijman I.J., Rodellar C., Valentini A., Williams J.L., Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus, Heredity 90 (2003) 350-358.

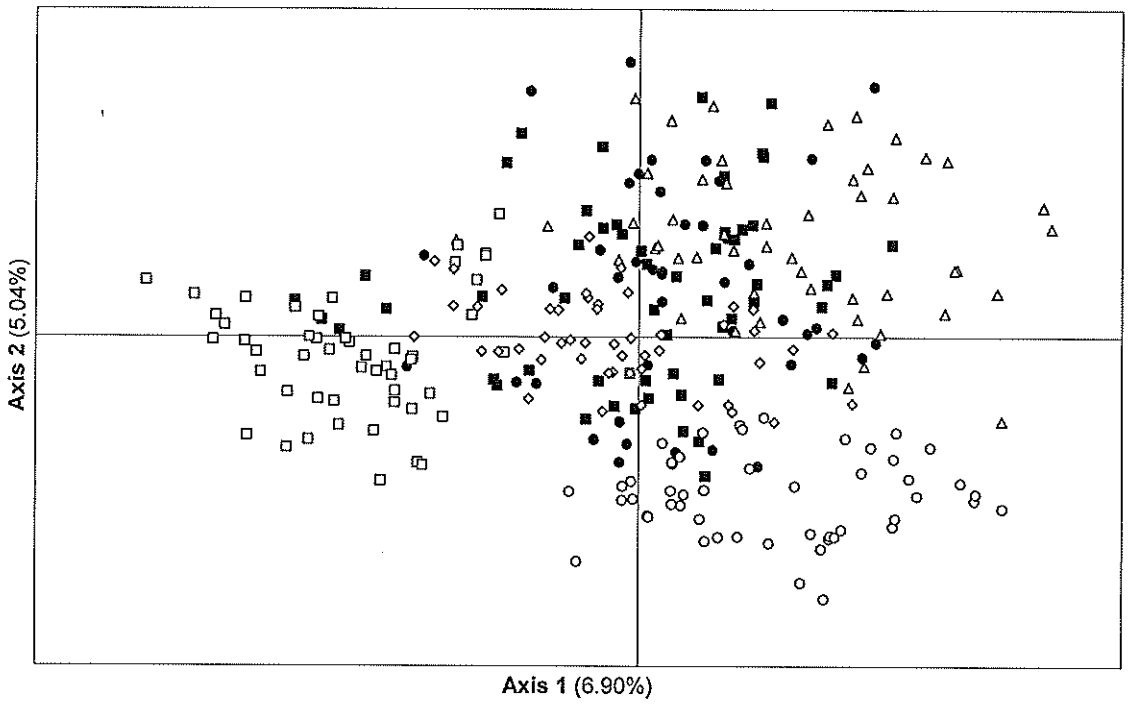
[47] Wright S., On the roles of directed and random changes in gene frequency in the genetics of populations, Evolution 2 (1948) 279-294.

Figure 1. Two-dimensional component analysis (PCA) on all individuals from generations G-2, G-1 and G11, using genotypes at all markers *(a)*, at markers in QTL regions *(b)* and at the supposedly neutral markers *(c)*. Black circles refer to G-2, black squares to G-1 and white items refer to G11: circles refer to Line 1, squares to Line 2, triangle to Line 3 and diamonds to the control line.

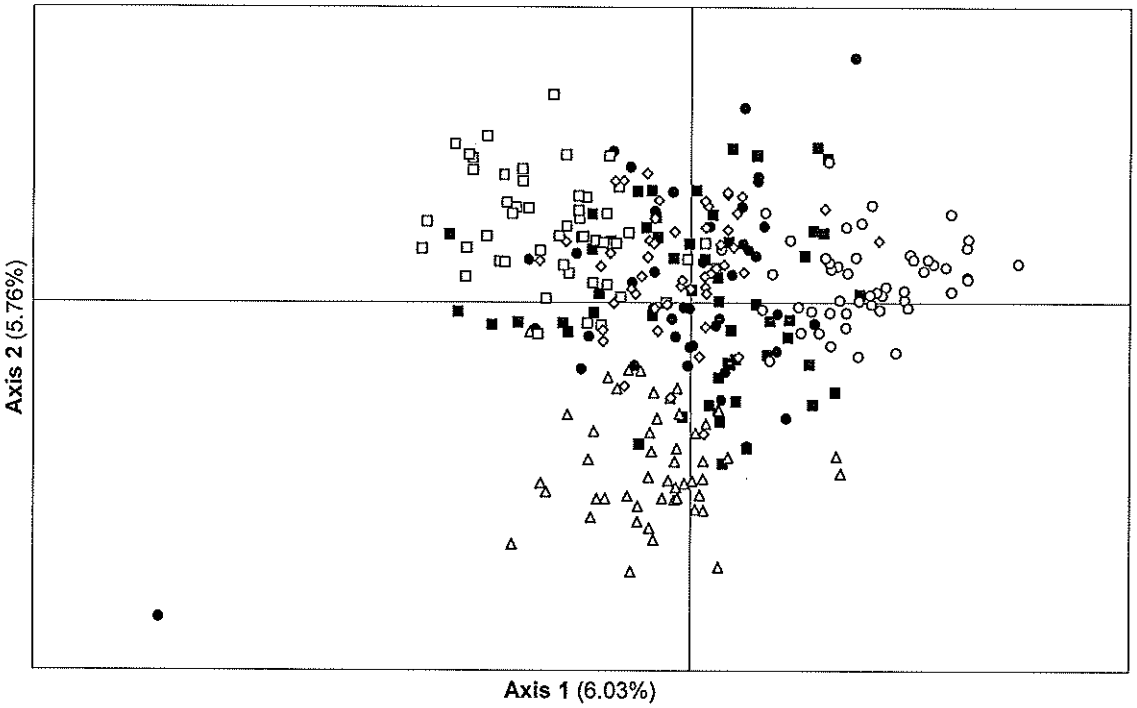
(a)



1 *(b)*



2
3 *(c)*



4

Table I. Position of markers in the QTL zones and the trait they are related to.

Zone	Marker	Chromosome	Position		Trait of QTL (Ab titre to...)
			Genetic (cM)	Physical (bp)	
1	*MCW183*	7	86	23417076	SRBC
	ADL279		92	24462410	
	ADL111		98	25777047	
	MCW236		109	28822966	
2	*ADL118*	14	0	2265471	KLH & *M. butyricum*
	MCW296		5	3665129	
	SEQALL0454		10	4774810	
	SEQALL0455		14	5695404	
	SEQALL0453		18	6830872	
3	*LEI146*	1	169	49939300	LPS
	ADL0359		172	52275623	
	SEQALL0426		191	57481907	
	SEQALL0427		192	57730587	
	SEQALL0428		195	58353741	
	MCW018		203	60171549	
	MCW112		205	61585157	
4	*ADL114*	2	319	111343871	SRBC
	LEI105		320	112311513	
	LEI355		325	112475918	
	SEQALL0433		335	115448137	
	GCT002		349	116794963	
	MCW166		360	124405931	
	MCW314		362	124918166	
5	*MCW306*	3	120	33953596	KLH
	ADL327		158	47104936	
6	*LEI166*	3	300	103360808	SRBC
	MCW037		317	106712843	
7	*LEI258*	16	0	147375	SRBC
	MCW370		0	160229	
	MCW371		0	158157	

Table II. Deviations from Hardy-Weinberg equilibrium as estimated by *Fis* values

Generation	Line	*LEI146*	*SEQALL427*	*MCW112*	*MCW306*	*ADL327*	*LEI258*	*MCW216*	*ADL278*
			[Zone 3]		*[Zone 5]*		*[Zone 7]*	*[Aviandiv]*	
G-2		-0.380[1]	-0.405[1]	-0.376[1]	-0.268[2]	-0.504[2]	-0.252[3]	-0.735[4]	-0.163
G11	**Line 1**	-0.156	0.408[1]	-0.117	0.070	0.358[2]	-0.063	-0.324	-0.143
	Line 2	-0.085	-0.146	-0.071	0.121	0.277[2]	0.054	-0.108	-0.326[4]
	Line 3	-0.235	0.202	0.055	0.018	-0.083	0.087	0.239	0.125
	Control	-0.001	0.142	-0.067	0.023	0.063	-0.024	0.104	-0.028

P-values: P =0.0071 *(1)*, P=0.025 *(2)*, P =0.017 *(3)*, and P =0.0023 *(4)*.

Table III. Estimated *Fst* values (and standard deviation) of markers involved in line differentiation.

Marker		*Fst* (± SD)
MCW183	*[Zone 1]*	0.196 (±0.133)
ADL111		0.158 (±0.056)
ADL118		0.188 (±0.053)
MCW296		0.163 (± 0.091)
SEQALL454	*[Zone 2]*	0.409 (± 0.108)
SEQALL455		0.336 (± 0.205)
SEQALL453		0.373 (± 0.170)
SEQALL426	*[Zone 3]*	0.236 (± 0.208)
MCW166	*[Zone 4]*	0.172 (± 0.122)
ADL327	*[Zone 5]*	0.217 (± 0.083)
LEI166	*[Zone 6]*	0.168 (± 0.146)
MCW370	*[Zone 7]*	0.223 (± 0.162)
ADL278		0.206 (± 0.091)
LEI234		0.178 (± 0.071)
MCW067	*[Aviandiv]*	0.189 (± 0.088)
MCW081		0.344 (± 0.178)
MCW222		0.158 (± 0.063)

Table IV. Genetic distances (± standard error) between the founder generation (G-2) and the four lines in generation G11. The upper matrix gives weighted Reynolds distances estimated by using genotypes of the supposedly neutral markers, whereas the lower matrix gives weighted Reynolds distances estimated by using genotypes of markers located in all QTL zones.

		G-2	G11			
			Control	Line 1	Line 2	Line 3
G-2		0	0.070 (±0.035)	0.095 (±0.031)	0.068 (±0.030)	0.069 (±0.027)
G11	Control	0.078 (±0.035)	0	0.106 (±0.030)	0.118 (±0.029)	0.070 (±0.028)
	Line 1	0.072 (±0.031)	0.153 (±0.030)	0	0.143 (±0.027)	0.147 (±0.025)
	Line 2	0.067 (±0.029)	0.127 (±0.029)	0.129 (±0.027)	0	0.121 (±0.025)
	Line 3	0.104 (±0.027)	0.148 (±0.028)	0.199 (±0.025)	0.176 (±0.025)	0

Table V. Observed allele frequencies of markers outside the 95% CI under the assumption of drift

Zone	Marker	Allele	Line 1 Obs.	Line 1 95% CI	Line 2 Obs.	Line 2 95% CI	Line 3 Obs.	Line 3 95% CI	Control line Obs.	Control line 95% CI
1	*MCW183*	292	0.170	*[0.085 ; 0.841]*	0.433	*[0.065 ; 0.768]*	0.540	*[0.120 ; 0.846]*	0.489	*[0.040 ; 0.799]*
		300	**0.702**	*[0 ; 0.584]*	0.4111	*[0 ; 0.445]*	0.070	*[0 ; 0.485]*	0.233	*[0 ; 0.651]*
		304	0.011	*[0 ; 0.607]*	0.156	*[0 ;0612]*	0.390	*[0 ; 0.538]*	0.244	*[0 ; 0.582]*
		308	0.117	*[0 ; 0.521]*	0	*[0 ;0.611]*	0	*[0 ; 0.549]*	0.033	*[0 ; 0.564]*
2	*ADL118*	156	0.650	*[0.222 ; 0.993]*	**0.260**	*[0.309 ;0.978]*	0.688	*[0.319 ; 0.969]*	0.622	*[0.283 ; 0.995]*
		157	0.350	*[0 ; 0.579]*	0.320	*[0 ;0.509]*	0	*[0 ; 0.523]*	0.378	*[0 ; 0.578]*
		160	0	*[0 ; 0.495]*	0.420	*[0 ;0.503]*	0.312	*[0 ; 0.481]*	0	*[0 ; 0.460]*
	SEQALL454	220	0.551	*[0 ; 0.576]*	0.398	*[0 ;0.582]*	0	*[0 ; 0.501]*	0.100	*[0 ; 0.582]*
		225	0.071	*[0 ; 0.708]*	0.561	*[0.003 ;0.654]*	0.110	*[0.011 ; 0.640]*	**0.900**	*[0.003 ; 0.737]*
		227	0.3673	*[0 ; 0.674]*	0.041	*[0 ;0.621]*	0.460	*[0.013 ; 0.664]*	0	*[0.009 ; 0.712]*
		229	0	*[0 ; 0.386]*	0	*[0 ;0.365]*	0	*[0 ; 0.424]*	0	*[0 ; 0.408]*
		231	0.010	*[0 ; 0.473]*	0	*[0 ;0.417]*	0.430	*[0 ; 0.434]*	0	*[0 ; 0.369]*
	SEQALL455	211	0.960	*[0.723 ; 1]*	**0.704**	*[0.748 ; 1]*	1	*[0.721 ; 1]*	1	*[0.7205 ; 1]*
		213	0.040	*[0 ; 0.277]*	**0.296**	*[0 ; 0.252]*	0	*[0 ; 0.215]*	0	*[0 ; 0.2795]*
	SEQALL453	203	0	*[0 ; 0.474]*	0.210	*[0 ; 0.403]*	0.051	*[0 ; 0.350]*	0	*[0 ; 0.354]*
		205	0.133	*[0 ; 0.429]*	0.120	*[0 ; 0.355]*	0	*[0 ; 0.429]*	0.042	*[0 ; 0.485]*
		209	0.041	*[0 ; 0.553]*	0.050	*[0.001 ; 0.663]*	0.296	*[0 ; 0.540]*	**0.750**	*[0 ; 0.589]*
		226	0.827	*[0.180 ; 0.939]*	0.620	*[0.163 ; 0.864]*	0.653	*[0.231 ; 0.933]*	0.208	*[0.164 ; 0.956]*
3	*SEQALL426*	153	0.867	*[0.091 ; 0.836]*	0.760	*[0.114 ; 0.821]*	**0.920**	*[0.093 ; 0.830]*	0.730	*[0.047 ; 0.838]*
		164	0.133	*[0.164 ; 0.909]*	0.240	*[0.179 ; 0.886]*	**0.08**	*[0.171 ; 0.907]*	0.270	*[0.162 ; 0.953]*

Table VI. Estimation of the effective population size for each line, using pedigree information (Ne_I) or genotype information (Ne_V) either from supposedly neutral markers or from markers in all QTL zones.

		Line 1	Line 2	Line 3	Control line
Ne_I		34	36	38	40
Ne_V	Neutral markers	43	46	48	56
[95% CI]		*[26 ; 66]*	*[28 ; 70]*	*[29 ; 74]*	*[34 ; 87]*
	Markers in all QTL zones	52	58	40	56
		[37 ; 81]	*[39 ; 85]*	*[27 ; 57]*	*[37 ; 81]*

1 ***Appendix 1:*** Position of the supposedly neutral markers from the Aviandiv panel.

Marker	Chromosome	Position (pb)
ADL268	1	82959056
LEI194	1	23950597
MCW111	1	35965715
LEI228	2	134463213
LEI234	2	10723773
MCW034	2	69660072
MCW206	2	30490372
MCW222	3	19350780
LEI094	4	50648081
MCW098	4	78890081
MCW295	4	16085321
MCW078	5	26438647
MCW081	5	45678311
LEI192	6	2416235
MCW014	6	6388946
ADL278	8	29243118
ADL112	10	20826244
MCW067	10	8887913
MCW216	13	11876856
MCW330	17	7009839
MCW069	26	1210140

Appendix 2: Observed allele frequencies for the markers located in the QTL zones

Zone	Marker	Allele	Generation G-2	Generation G11			
				Line1	Line2	Line3	Line4
1	*MCW183*	292	0.470	0.170	0.433	0.540	0.489
		300	0.169	0.702	0.411	0.070	0.233
		304	0.157	0.011	0.156	0.390	0.244
		308	0.205	0.117	0	0	0.033
	ADL279	101	0.025	0	0	0.041	0
		111	0.848	0.840	0.680	0.745	0.827
		113	0.076	0.160	0.320	0.214	0.174
	ADL111	99	0.051	0	0	0	0
		123	0.361	0.380	0.020	0.071	0.250
		124	0.374	0.400	0.730	0.337	0.630
		134	0.265	0.220	0.250	0.592	0.120
	MCW236	308	0.073	0	0	0.276	0.260
		311	0.659	1	0.730	0.551	0.660
		313	0.049	0	0.240	0	0.030
		315	0.146	0	0.030	0.092	0.050
		323	0.073	0	0	0.082	0
2	*ADL118*	156	0.561	0.650	0.260	0.688	0.622
		157	0.281	0.350	0.320	0.313	0.378
		160	0.159	0	0.420	0.313	0
	MCW296	238	0.161	0.350	0.320	0	0.061
		242	0.840	0.650	0.680	1	0.939
	SEQALL0454	220	0.386	0.551	0.398	0	0.100
		225	0.229	0.071	0.561	0.110	0.900
		227	0.205	0.367	0.041	0.460	0
		229	0.060	0	0	0	0
	SEQALL0455	231	0.121	0.010	0	0.430	0
		211	0.962	0.960	0.704	1	1
		213	0.039	0.040	0.296	0	0
	SEQALL0453	203	0.073	0	0.210	0.051	0
		205	0.134	0.133	0.120	0	0.042
		209	0.195	0.041	0.050	0.296	0.750
		226	0.598	0.827	0.620	0.653	0.208

Zone	Marker	Allele	Generation G-2	Generation G11			
				Line1	Line2	Line3	Line4
3	*LEI146*	252	0.136	0	0.170	0	0
		262	0.012	0	0	0	0
		263	0.407	0.220	0.270	0.470	0.650
		274	0.444	0.780	0.560	0.530	0.350
	ADL0359	204	0.769	0.867	0.760	0.920	0.730
		216	0.231	0.133	0.240	0.080	0.270
	SEQALL0426	153	0.469	0.380	0.520	0.030	0.580
		164	0.531	0.620	0.480	0.970	0.420
	SEQALL0427	178	0.657	0.585	0.778	0.704	0.740
		180	0.012	0.011	0	0.031	0.010
		184	0.060	0.351	0	0.214	0.180
	SEQALL0428	186	0.277	0.053	0.222	0.051	0.070
		221	0.309	0.082	0.146	0.150	0.320
		223	0.691	0.918	0.854	0.850	0.680
	MCW018	221	0.247	0	0.150	0.060	0.050
		227	0.753	1	0.850	0.940	0.950
	MCW112	258	0.256	0.460	0.140	0.480	0.430
		271	0.671	0.540	0.790	0.3470	0.560
		273	0.073	0	0.070	0.174	0.010
4	*ADL114*	163	0.675	0.744	0.600	0.340	0.710
		176	0.013	0	0	0	0
		179	0.313	0.256	0.400	0.660	0.290
	LEI105	134	0.269	0.250	0.071	0.070	0.220
		142	0.089	0	0	0	0
		147	0.385	0.531	0.622	0.290	0.430
		151	0.256	0.219	0.306	0.640	0.350
	LEI355	223	0.024	0	0	0	0
		227	0.366	0.300	0.070	0.071	0.122
		235	0.329	0.210	0.260	0.633	0.449
		239	0.244	0.440	0.6200	0.2959	0.4286
		243	0	0	0.0200	0	0
		247	0.036	0.050	0.030	0	0
	SEQALL0433	232	0.096	0	0.010	0.080	0.040
		236	0.337	0.250	0.540	0.190	0.110
		237	0.566	0.750	0.450	0.730	0.850
	GCT002	151	0.756	0.980	0.900	0.725	0.910
		161	0.122	0.020	0.100	0	0.050
		167	0.122	0	0	0.276	0.040
	MCW166	191	0.072	0.160	0	0.225	0.030
		193	0.482	0.250	0.680	0.306	0.170
		201	0.446	0.590	0.320	0.469	0.800
	MCW314	273	0.232	0.230	0.370	0.360	0.190
		277	0.768	0.770	0.630	0.640	0.810

Zone	Marker	Allele	Generation G-2	Generation G11 Line1	Line2	Line3	Line4
5	*MCW306*	146	0.072	0	0	0	0
		148	0.482	0.480	0.460	0.280	0.370
		160	0.374	0.520	0.520	0.720	0.630
		162	0.072	0	0.020	0	0
	ADL327	107	0.241	0.622	0.204	0.100	0.200
		120	0.217	0	0.378	0.080	0.070
		99	0.542	0.378	0.418	0.820	0.730
6	*LEI166*	251	0.293	0.021	0.464	0.426	0.289
		255	0.024	0.021	0.060	0	0
	MCW037	261	0.683	0.957	0.476	0.575	0.711
		150	0.098	0.235	0.177	0.041	0.090
		152	0.476	0.480	0.646	0.643	0.470
		153	0	0	0	0	0.030
		154	0.427	0.286	0.177	0.316	0.410
7	*LEI258*	261	0.220	0.180	0.360	0.480	0.388
		359	0.366	0.280	0.050	0.408	0.225
		422	0.012	0.130	0	0	0.031
		489	0.256	0.410	0.450	0.051	0.357
		544	0.146	0	0.130	0.061	0
		557	0	0	0.010	0	0
	MCW370	178	0.173	0.010	0.117	0.061	0
		179	0.2716	0.4271	0.4681	0.0510	0.370
		181	0.383	0.375	0.064	0.418	0.250
		AAA	0.173	0.188	0.351	0.469	0.380
	MCW371	201	0.222	0.174	0.361	0.539	0.409
		202	0	0.012	0	0	0
		203	0.778	0.814	0.640	0.462	0.591

Appendix 3: Observed allele frequencies for the supposedly neutral markers (excluding those located in a QTL zone)

Marker	Allele	Generation	Generation G11			
		G-2	Line1	Line2	Line3	Line4
ADL112	123	0.051	0.030	0.106	0	0.052
	127	0.949	0.970	0.895	1	0.948
ADL268	109	0.696	0.500	0.640	0.745	0.640
	111	0.266	0.150	0.160	0.194	0.100
	113	0.038	0.350	0.200	0.061	0.260
ADL278	110	0.188	0.133	0.160	0	0.460
	116	0.063	0	0	0.110	0
	117	0.200	0	0.360	0.490	0.090
	119	0.550	0.867	0.480	0.400	0.450
LEI094	246	0.085	0	0.367	0.070	0.089
	260	0.476	0.628	0.389	0.540	0.456
	262	0.256	0.096	0.244	0.330	0.278
	283	0.183	0.277	0	0.060	0.178
LEI192	253	0.573	0.710	0.370	0.880	0.700
	265	0.207	0.100	0.480	0.120	0.210
	269	0.183	0.080	0.040	0	0.020
	572	0.037	0.110	0.110	0	0.070
LEI194	128	0.104	0	0	0	0
	130	0.571	0.600	0.598	0.850	0.844
	138	0.182	0.350	0.402	0	0.156
	160	0.013	0	0	0.110	0
	164	0.117	0.050	0	0.040	0
	172	0.013	0	0	0	0
LEI228	197	0.750	0.640	0.783	0.771	0.735
	208	0.138	0.040	0.022	0.073	0.184
	216	0.025	0	0	0	0
	224	0.075	0.320	0.185	0.021	0.082
	227	0	0	0	0.073	0
	235	0.013	0	0.011	0.063	0
LEI234	215	0.167	0.080	0.096	0.020	0.260
	280	0.060	0.210	0.032	0	0.073
	288	0.107	0.280	0	0.092	0
	297	0.655	0.430	0.872	0.878	0.667
	305	0.012	0	0	0.010	0
MCW014	162	0.063	0	0	0	0
	176	0.813	1	1	0.980	0.896
	178	0.125	0	0	0.020	0.104
MCW034	219	0.305	0.080	0.280	0.122	0.120
	230	0.183	0.220	0.280	0.133	0.030
	231	0.512	0.700	0.440	0.745	0.850
MCW067	174	0.217	0.180	0.280	0.420	0.180
	176	0.361	0.260	0	0.390	0.620
	180	0.361	0.450	0.240	0.190	0.200
	182	0.060	0.110	0.480	0	0
MCW069	154	0.434	0.830	0.790	0.660	0.790
	161	0.566	0.170	0.210	0.340	0.210

Marker	Allele	Generation G-2	Generation G11 Line1	Line2	Line3	Line4
MCW078	134	0.386	0.490	0.500	0.190	0.222
	138	0.554	0.510	0.500	0.780	0.778
	140	0.060	0	0	0.030	0
MCW081	111	0.410	0	0.321	0.052	0.022
	131	0.590	1	0.679	0.948	0.978
MCW098	255	0.241	0.020	0.032	0.194	0.250
	257	0.760	0.980	0.968	0.806	0.750
MCW111	97	0.063	0	0.122	0.160	0.244
	99	0.938	1	0.878	0.840	0.756
MCW206	223	0.163	0.440	0	0.040	0.150
	228	0.663	0.370	0.700	0.650	0.480
	236	0.175	0.190	0.300	0.310	0.370
MCW216	142	0.456	0.250	0.255	0.417	0.406
	144	0.544	0.750	0.745	0.583	0.594
MCW222	218	0.463	0.404	0.567	0.510	0.333
	220	0.207	0	0.389	0.060	0.400
	222	0.329	0.596	0.044	0.430	0.267
MCW295	83	0.301	0.380	0.255	0.061	0.146
	85	0.145	0	0.213	0.398	0
	93	0.108	0.180	0.138	0.184	0.219
	95	0.121	0	0.149	0.225	0.260
	97	0.325	0.440	0.245	0.133	0.375
MCW330	274	0.800	0.862	0.881	0.745	0.867
	286	0.200	0.138	0.119	0.255	0.133

1

2

CHAPTER 5: Evaluation of risks consequently to reduction of genetic diversity in populations having experimented a strong bottleneck

This chapter deals with consequences of intensive selection on the increasing number of genetic abnormalities in simulated populations. We focus on dairy cattle populations, mainly on the French Prim'Holstein population since those populations have faced severe bottlenecks. We aim at mimicking mechanism of the reproductive system of this population in order to model and simulate polymorphism evolution of specific alleles: deleterious alleles linked to genetic defects. Appearance of those alleles is the direct consequences of diversity loss in the population; then, management of deleterious alleles is part of management of genetic diversity. But efficient management requires knowledge of the genetic basis.

The goal of this chapter is to improve our knowledge about appearance of genetic defects in selected populations and the consequences of counter-selection against the deleterious alleles in short and long-term.

More or less empirical selection procedures have been going on in animal populations for thousands of years. During the last two centuries, breeds have been standardised and, from the middle of the 20th century, breeding programmes applying the quantitative genetics concepts and methods have been implemented. In the last decades the selection procedures became more and more accurate and intense. Such selection, over many generations and in large populations, has driven the accumulation of mutations with favourable effects on phenotypes, as well as mutations with pathological phenotypic consequences and deleterious mutations.

In the case of bovine populations, several morphological, neurological or metabolism abnormalities have been reported over the last ten years. For about 400 of these abnormalities, a genetic basis is putative or has been proven; for 56 of them, a locus was located and for 27 of them, a causal mutation was identified (Ducos *et al*., 2004). In France, a large scale inventory has been organized by the so called "French Bovine Genetic Disease Observatory", created in 2002. The incidence of genetic defects is usually low but economical consequences of those abnormalities are important. Vigil programmes were set up in order to detect any fast increase of the frequency of some particular defects and manage the crises. When then number of stillbirth calving peaks, counter-selection against the defect is handled by developing a molecular test to find the causal mutation out and by selecting against the mutant allele in the breeding scheme.

In dairy cattle breeding, the massive and unbalanced use of some artificial insemination (AI) bulls induces large bottlenecks from one generation to the next one. As a consequence, many "large" dairy cattle breeds, i.e. breeds widely used, with a very large female population size, are small populations from a genetic point of view (for recent results in the case of French breeds, see Mattalia *et al.*, 2006).

The goal of this study is to investigate the risk of appearance of genetic defects in selected populations submitted to bottlenecks, as dairy cattle populations. Is the increase of abnormalities due to new mutant alleles or due to the load of deleterious alleles inherited from the past? Is the probability of facing peaks due to the reduction of effective size of population or due to the specific selection scheme? What would be the most efficient management against a new peak: temporary or continual selection against the deleterious allele?

1. Material and methods

The evolution of the frequency of a deleterious allele was monitored within a simulated population. In order to mimic the intensification of the selection procedures having occurred from the 1950s', two successive periods were considered (Figure 4): during the first period, natural service bulls were used, there was no artificial selection and, within a given parental sex, no distinction was made between breeding animals according to the sex of their future progeny; during the second period, the number of sires was very small, the selection intensity was as high as possible and the fourth gene pathways (sire-sire, dam-sire, sire-dam, dam-dam) were distinguished. The demographic parameters of the simulated population were inspired from the current status and the known history of the French Holstein population. However, for the sake of simplicity, the population was simulated with separate generations.

For most abnormalities, a Mendelian inheritance, mainly autosomal recessive (60%), has been postulated (Huston, 1993). Therefore, a locus with two alleles was considered in the simulation: a wild-type allele $\mathcal{M}$, that mutates at rate u to a deleterious allele m (no back mutations), with a mutation rate ranging between 10^{-4} and 10^{-8} from one studied situation to the other. The deleterious allele was assumed to be recessive and lethal: the selective value (w_i) of an individual i was equal to 1 for $\mathcal{MM}$ homozygote and heterozygote $\mathcal{M}m$ individuals, whereas w_i was equal to 0 for mm homozygote individuals. For this study, the BLAD abnormality (Bovine Leukocyte Adhesion Deficiency) was considered as a reference frame: the occurrence rate of BLAD at birth was estimated to be 0.2% in the United States in 1992, which corresponds to a 0.044 frequency of the lethal allele in the population. Frequency of the causal mutant allele (D128G allele) was estimated to be 0.15 and 0.06 for Holstein bulls and dams respectively (Nagahata, 2004). In Denmark, of the tested animals in 1993, 0.5% of them were affected animals. As BLAD in different Holstein populations has been successfully controlled, the above frequencies will define a peak of abnormality in the population: the population will be considered to face a peak of abnormality when frequency of the lethal allele reaches 0.044, which corresponds to a 0.2% occurrence rate of stillbirth. Furthermore, we will considerer a 10^{-4} mutation rate.

1.1. Initialisation

The purpose of the initialisation step is to estimate the frequency of the deleterious allele when (intense) selection starts, i.e. at the end of the first period (see above). During this first period, the population was assumed to be unselected (random choice of parents). It counted 50,000 dams in each generation, and one sire for 30 to 50 dams. Mating was at random. Under such hypotheses, the theoretical value of the effective size (*Ne*), is given by the classical equation (e.g. Crow and Kimura, 1970, p. 103): $1/Ne = 1/4N_m + 1/4N_f$, where N_m and N_f are the number of sires and dams, respectively. Then, the effective size of the population was comprised between 4,000 and 6,000, according to the extreme values for the number of dams per sire. The population is supposed to have reached the mutation-selection-drift equilibrium by that time; therefore, frequency of the inherited deleterious allele m at the end of the first period could be estimated through diffusion equations (Crow and Kimura, 1970, chapter 8). Then, allele frequency in the base population was estimated to be between 10^{-2} and 10^{-4}. Furthermore, we will consider q_0 to take 10^{-2} value.

1.2. Transmission: reproduction and selection during the second period

According to demographic data from the French herd book of Holstein population, the population counted 4,150 males and 1,900,000 females. In total 2,700 sires and 424,000 dams were referenced, but among them, only 275 sires and 1,365 dams were considered as the elite because of their high breeding values. Those referenced figures will be used in our simulated population and the "elite" reproducers will procreate the next generation of bulls: 275 different bull-sires and 1,365 bull-dams will be used (see Fig.4). Furthermore, those figures will be referred as "demographic data".
Mating was at random between individuals selected as parents to procreate the next generation.

In a first set of simulations (hereafter referred as "random choice") no artificial selection of the parents occurred, so that the population was just submitted to the selection on the $\mathcal{M}/m$ locus (see next).

In a second set of populations (thereafter referred as “mass selection”), parents were selected according to their phenotypic value for a given trait. The underlying additive model to generate the phenotypic value P_i of an individual i was $P_i = \mu + A_i + E_i$, where μ is the overall mean, A_i is the individual genetic additive value and E_i is the environmental effect. For the founder animals, breeding values (A_i) were generated as $A_i = \sigma_A \cdot \varepsilon_i$, where ε_i is a random number, normally distributed, centred on zero and independent from one founder animal to the other, and σ_A is the additive genetic standard deviation in the base population. On the same way, environmental effect was generated as $E_i = \sigma_E \cdot \gamma_i$, where γ_i is a random number, normaly distributed, centred on zero, independent from ε_i and independent from one founder animal to the other, and σ_E is the environmental standard deviation, assumed to be constant over generations. From one generation to the other, the transmission model was as follows: for each offspring, i, $A_i = \frac{1}{2} A_p + \frac{1}{2} A_m + Z_i$, where A_p et A_m are the breeding values of its sire and the dam, respectively, and Z_i represents the meiosis alea. This alea was generated as follows: $Z_i = \sigma_A \cdot \zeta_i \cdot \left[\frac{1}{2} \left(1 - \frac{F_p + F_m}{2} \right) \right]^{1/2}$, where ζ_i is an independent random number drawn from a N(0,1) distribution, and F_p and F_m are the coefficients of inbreeding of the sire (F_p) and the dam (F_m), respectively (Foulley and Chevalet, 1971; Verrier et al., 1989). The individual coefficients of inbreeding were computed on the basis of the complete knowledge of the pedigree. The additive genetic standard deviation was arbitrarily chosen as unity in the base population ($\sigma_A = 1$) and the heritability of the selected trait was 0.3.

At each generation, the genotype of animals at the locus $\mathcal{M}/m$ was taken into account. First, mm homozygous animals were systematically considered as dead at birth. Second, when the frequency of the lethal allele (m) was higher than a given threshold, fixed in this study to 0.044 (see above), a systematic selection against this allele was assumed, by putting to zero the selective value (w_i) of $\mathcal{M}m$ homozygous animals. This corresponds to the design of molecular tests allowing the systematic detection and the elimination of individual carrying the lethal allele. This new selection regime could either be continual or temporary, *i.e.* counter-selection of heterozygote $\mathcal{M}m$ stopped as soon as the frequency of m was lower than the threshold.

Simulations (1000 iterations) starting from the base population over *t* generations were undergone. Frequency of m allele (q_t) was given by $q_t = q_0 + q'$, where q_0 is the inherited allele frequency (see *Initialisation*) and *q'* the frequency of a new allele appeared over *t* generations.

1.3. Estimation of the variance effective size of the population

A locus, independent from both the M/m locus and the selected trait, and with no mutation, was simulated. This locus was biallelic, with the same initial frequencies (in the base population) as for the M/m locus. The "variance" effective size of the population (Ne_V) was estimated on the basis of the temporal variation of the allele frequencies at the neutral locus (Waples, 1989), using estimator proposed by Nei and Tajima (1981).

1.4. Duration of the process and number of replicates

Simulations over 10 generations were undergone with random choice of the parents, in order to understand the underlying dynamics of new mutant allele appearance. The 10-generations duration matches with the stretch of time starting from 1950 up till nowadays, considering that the generation interval for French Holstein population is estimated to be 5.7 years (Mattalia *et al.*, 2006). In the situation of selection, simulations were run up to 30 generations. For each case, 1000 replicates were undergone.

1.5. Influence of the selection scheme and the reduction of the effective size of population

Studies from Institut de l'Elevage have shown that 100 bull-sires were active nowadays but only 10 of them procreate 80% of the young bulls, that the ratio bull-sires per bull-dams was 1/5 and that bull-dams represented 1‰ of the total number of dams. In order to define the useful number of reproducers, we reduced the number of sires and dams: the different hypotheses that were tested are presented in Table 4.

In order to test the influence of the complex selection scheme, we have simulated a population with the same number of sires but undergoing the 'natural' animal breeding scheme handled before 1950's (in average, 40 dams per bull). Dams and sires were selected according to their

phenotypic value for a given trait ("mass selection"). The number of sires and dams to be mated depends of the useful number of reproducers defined as above.

1.6. Risk of facing new abnormalities

In order to evaluate the risk of facing a peak at another locus, i.e. facing a new abnormality, when counter-selection occurred against the lethal allele (m), we simulated other loci carrying lethal alleles.

The loci that would carry a lethal allele may be considered as lethal equivalents. Their number (1000) was estimated from the concept of expected number of lethal equivalents (n_{le}): $n_{le} \approx 2\bar{q}n$ (Hedrick *et al.*, 1998), where $\bar{q}$ is the average frequency of lethal alleles and n is the number of loci in an organism that can carry a lethal allele. In the bovine genome, there are 30 000 genes in average; assuming that a locus does not match directly with a gene, then the total number of loci should be larger, and assuming that $\bar{q}$ was approximated by the inherited allele frequency q_0, we have decided to considerer 1000 lethal equivalents in our population, *i.e.* 1000 independent loci carrying lethal alleles.

2. Results

2.1. Parents chosen at random

After 10 generations, the average frequency q_t was almost equal to q_0, frequency of a new allele q' being negligible (average q' was 10^{-3}). The average q_t was equal to 9.1 10^{-3} in each gene pool (sire, dam or whole population).

Persistence of a new mutant allele differed according to the sub-population where it appeared. In the dam population, new mutant alleles that were still observed after 10 generations had appeared mostly in the very last generations (10 % in generation 8, 22% in generation 9 and 57% in generation 10), whereas, in the sire population, new mutant allele had appeared in any of the 10 generations, with the same probability. Then, we may conclude that the increase of abnormalities was due to the load of deleterious alleles inherited from the past, rather than new mutant alleles.

2.2. Mass selected parents

a) Short-term period

From the simulated replicates, a 95% confidence interval (CI) for the final frequency of the mutant allele was drawn in each gene pool (sire, dam and offspring pools). The 95% CI under different hypotheses are presented in Table 5.

Observed allele frequencies using demographic data for dams and sires fell outside the 95% CI for each gene pool.
The reduction of the number of bull-dams did not have influence on CI bounds.
When the number of sires was reduced from 2,700 to 100, of which 10 bull sires, and the number of dams fitting demographic data, observed frequency of the deleterious allele in the whole population (0.044) matched with the 95% CI, but CI bounds were not large enough for the sire and dam gene pools. The observed frequency of the deleterious allele in the whole population matched with the 95% CI when the number of sires was reduced to 25, of which 5 bull sires, for each gene pool.
Out of the 1000 simulated population frequencies after 10 generations, 55 of them were larger than the observed frequency of the deleterious allele when total number of sires was 100, of which 10 bull sires (55 in the case of 25 sires in total, of which 25 bull sires). This means that in average, among 1000 independent loci, 55 of them showed an increase of frequency (frequency higher than the threshold defined to start a selection against the deleterious allele) that may lead to an abnormality peak; then, we could estimate risk of facing a new crisis at 5.5%.

For each gene pool, 95% CIs were similar, whatever the selection scheme (‘natural’ or complex), which lead us to conclude that abnormalities peaks are mainly due to the very small number of reproducers, more than the specific selection scheme.

b) Long-term period

Simulations over 30 generations were undergone with mass selection of the parents in the population in order to get a long-term overview on mutant allele appearance.

Figure 5 shows evolution of average lethal allele frequency and its 95% CI when continual or temporary counter-selection against the lethal allele occurred in a population with 25 sires, of which 5 bull-sires. Counter-selection started in generation 2, which is illustrated by the deleterious allele threshold line (lethal allele frequency equals to 0.044) crosses the 95% CI. Average frequency rapidly decreased and stabilised after 15 generations at around 8.10^{-4}.when total number of sires was 25, of which 5 bull-sires.
When increasing the number of sires with total number of sires was 100, of which 10 bull-sires, curve of evolution of average lethal allele frequency overlapped (results not shown) whatever the counter-selection regime.

Figure 6 shows the risk of facing a peak (number of loci out of 100 with a frequency higher than deleterious allele threshold) over 30 generations when continual or temporary counter-selection against the lethal allele occurred. The risk was higher during the first 10 generations when total number of sires was 25, whatever counter-selection against the lethal allele was temporary or continual. From 10 to 30 generations, the risk was equivalent whatever the number of sires of method of counter-selection against the lethal allele and was estimated in average at 0.2%.
Figure 7 shows the risk of facing a peak at another locus over 30 generations when temporary or continual counter-selection occurred against the lethal allele.
When temporal counter-selection occurred, the risk was higher when total number of sire was 100, of which 10 bull sires, than when total number of sires was 25, of which 5 bull sires. On the contrary, when continual counter-selection occurred, risk was lower when total number of sire was 100, of which 10 bull sires, than when total number of sires was 25, of which 5 bull sires. In the long term, whatever the total number of sires or the method for counter-selection against lethal alleles, average risk stabilised at 1%. This means that among 1000 independent loci, 10 of them shows an increase of frequency (frequency higher than BLAD threshold of 0.044) that may lead to an abnormality peak.

Table 6 shows estimations of the “variance” effective population size and inbreeding after 10 generations. Estimations of the effective population size and inbreeding rate were similar with the complex selection scheme with 25 sires, of which 5 bull-sires, and with the ‘natural’ selection scheme with 25 sires and 1000 dams. When the total number of sires increased from 25 to 100, the “variance” effective size of the population slightly increased but the inbreeding rate was divided by half.

3. Discussion

Deleterious alleles regularly appear and frequency of some of them may increase in the dam gene pool but strong bottlenecks from one generation to another one act like a purge. As a matter of fact, unbalanced use of sires is the cause of abnormalities crisis. In fact, alleles with nowadays fast frequency increase are alleles inherited and accumulated along domestication years, before intensification of selection in the 1950's; and an intense use of some male reproducers allows mutation load (Glemin, 2003) to express. If one of the highly-used sires carries a lethal allele, the risk of a fast frequency increase of this allele is strong. It is actually what happened for BLAD crisis: BLAD carriers were among the most prominent bulls of Holstein breed, such as Osbornedal Ivanhoe, worldwide-used bulls, and some of its sons.

Mechanism of French Holstein population could actually be compared to mechanism of a population where only 25 sires procreate, assuming very simplified hypotheses such as non-overlapping generations and simple mass selection. In fact, the intense artificial selection undertaken nowadays is much more accurate than mass selection. Genetic improvement is greatest through selection of males because fewer males than females are needed in breeding: bull-sires and bull-dams pathway contribute around 30% and 39%, respectively, to the overall genetic improvement. Progeny testing of bulls has become central to genetic improvement of dairy cattle and the intense use of some bulls require an accurate estimation of their breeding values (high precision of their indexes with an accuracy of the estimation at least greater than 0.7). Thus, consequences on genetic diversity that were drawn out of this study actually minimize the real effect of selection and then, minimize risks. On the contrary, consequences of overlapping generations are more difficult to evaluate but we may consider that overlapping generations intensify the effect of selection since reproducers of different generations may compete against one another at the same time.

Moreover we assume fitness to be linear, giving every parent the same probability to procreate; in reality, fitness should be exponential since it is defined according to breeding value, decreasing again the number of useful parents. Then, estimation of effective size would not be large and would agree with estimation (Ne_I = 50) based on pedigree analysis (Boichard *et al*, 1997), although estimation of effective population size is always lower when using pedigree data, compared to estimation based on allele variation (Ne_V). Finally, incidence of the lethal allele (0.2%) was calculated as q^2, assuming non-inbred population, so that frequency of the lethal allele was estimated to be 0.044 in the population; however, according

to inbreeding estimation, incidence of the lethal allele should take inbreeding into account and be calculated as $q^2 + Fq(1-q)$, so that frequency of the lethal allele would be lower (0.017) in the population.

This study illustrates the visible consequence of inbreeding, which increases with the reduction of the effective size of the population. Inbreeding in itself simply increases homozygosity, whether homozygous combinations contain deleterious allele or not; therefore, it increases the chances of deleterious allele, particularly lethal genes, to become homozygous and thus express themselves.

Continual counter-selection against lethal allele that currently occurs, such as against BLAD allele, will not have direct consequences on frequency of other deleterious alleles. However, probability of facing a new crisis at each mutant locus still exist (1%, *i.e.* 1 locus out 100). An increase of the number of reproducers (as for instance, 100 sires, of which 10 bull-sires) would correspond to a better use of reproducers, would not have any impact on the risk for other loci in the Holstein population but would at least reduce inbreeding coefficient and thus, inbreeding depression.

Numerous abnormalities are well known in the French Holstein population, such as umbilical hernia. Frequency of those abnormalities in the population is still very low. But would happen if a new sire with high breeding value was carrier of one of those alleles? What would happen if one son of a worldwide used Holstein bull would receive a lethal allele from his mother and then would be intensively used?

In the light of this study, advice could be given for the management of population which selection tends to intensify but have not faced any abnormalities peaks yet, especially for populations with high-level diffusion possibility of parents, sheep populations for instance:

- The number of reproducers should not be too narrow and their use should be balanced, especially for sires.
- If frequency of a lethal allele ever increases in the population, temporary counter-selection against the lethal allele should be preferred in order to low risk at other loci.

Table 4: Different hypotheses about the reduction of the number of reproducers.

Dams		**Sires**	
Total number	***Number of bull-dams***	***Total number***	***Number of bull-sires***
125,000	125	100	10
424,000	1365	100	10
(Demographic data)		25	5

Table 5: 95% confidence intervals for the final frequency q_t of the mutant allele after 10 generations with mass selection of the parents

		Whole population	**Sire gene pool**	**Dam gene pool**
Complex selection scheme	*Demographic data*	[0.004; 0.019]	[0.003; 0.020]	[0.004; 0.017]
	Reduction of number of sires			
	100 sires of which 10 bull-sires	[0; 0.062]	[0; 0.075]	[0; 0.056]
	25 sires of which 5 bull-sires	[0; 0.082]	[0; 0.010]	[0; 0.083]
'Natural' selection scheme	25 sires & 1000 dams	[0; 0.082]	[0; 0.080]	[0; 0.080]

Table 6: Estimation of the effective population size and inbreeding after 10 generations

		Choice of parents	Ne_V	***F*** (%)
Complex selection scheme	Demographic data	*mass selection*	2868	0.17
	100 sires, of which 10 bull-sires		240	5.40
	25 sires, of which 5 bull-sires		165	10.66
'Natural' selection scheme	25 sires & 1000 dams	*mass selection*	158	11.19
		random	190	10.63

Figure 4: Schematic representation of the selective scheme in dairy cattle: *(a)* before 1950's ('natural' breeding scheme), and *(b)* after 1950's (complex selective scheme).

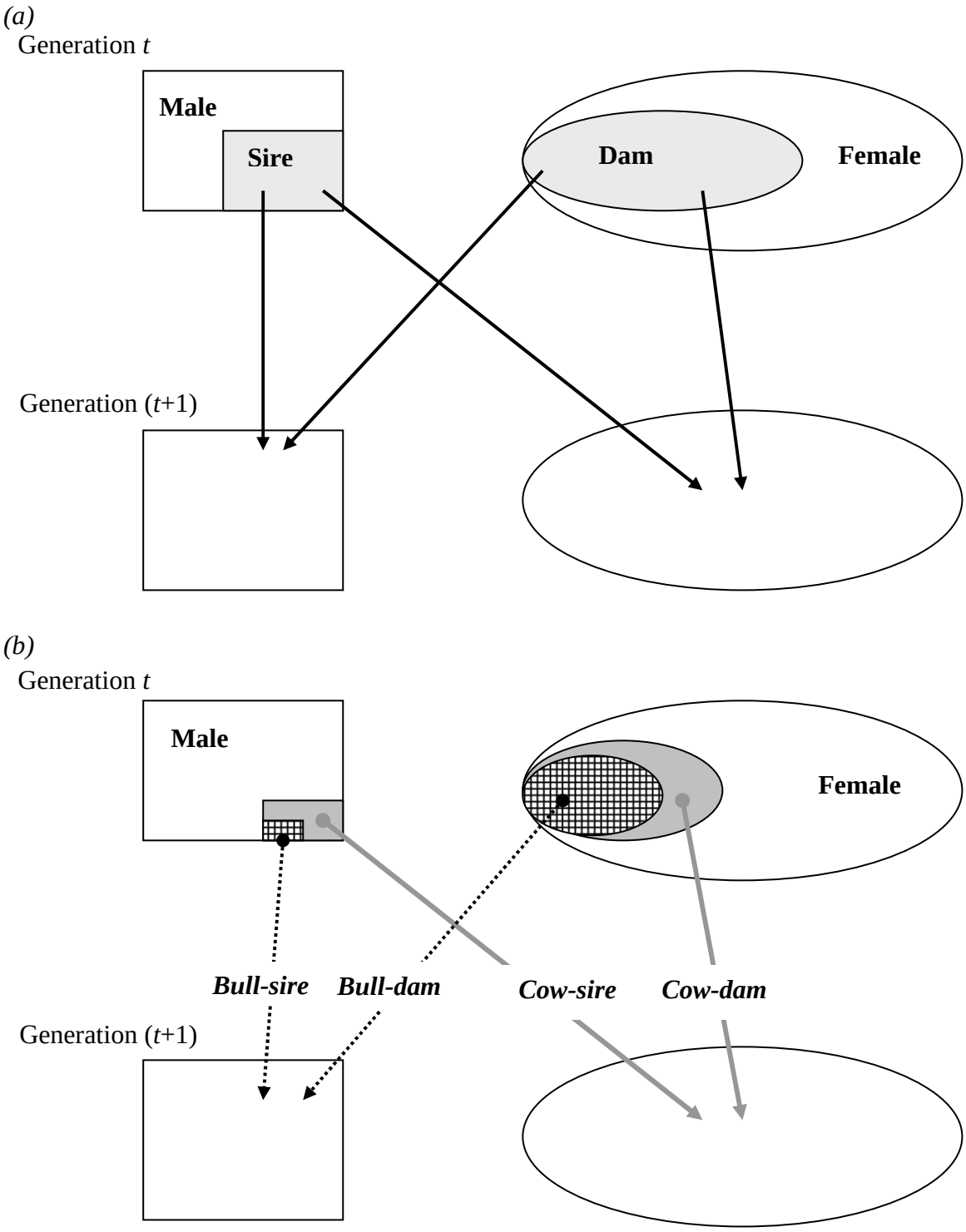


Figure 5: Evolution of the average allele frequency in the whole population over 30 generations when continual (triangle) or temporal (circle) counter-selection against the lethal allele occurred, assuming total number of sires to be 25, of which 5 bull-sires. Dotted lines refer to bounds of the 95% CIs: bold dotted lines for temporal counter-selection and plain dotted line for continual counter-selection. Semi-dotted line refers to the BLAD threshold lethal allele frequency.

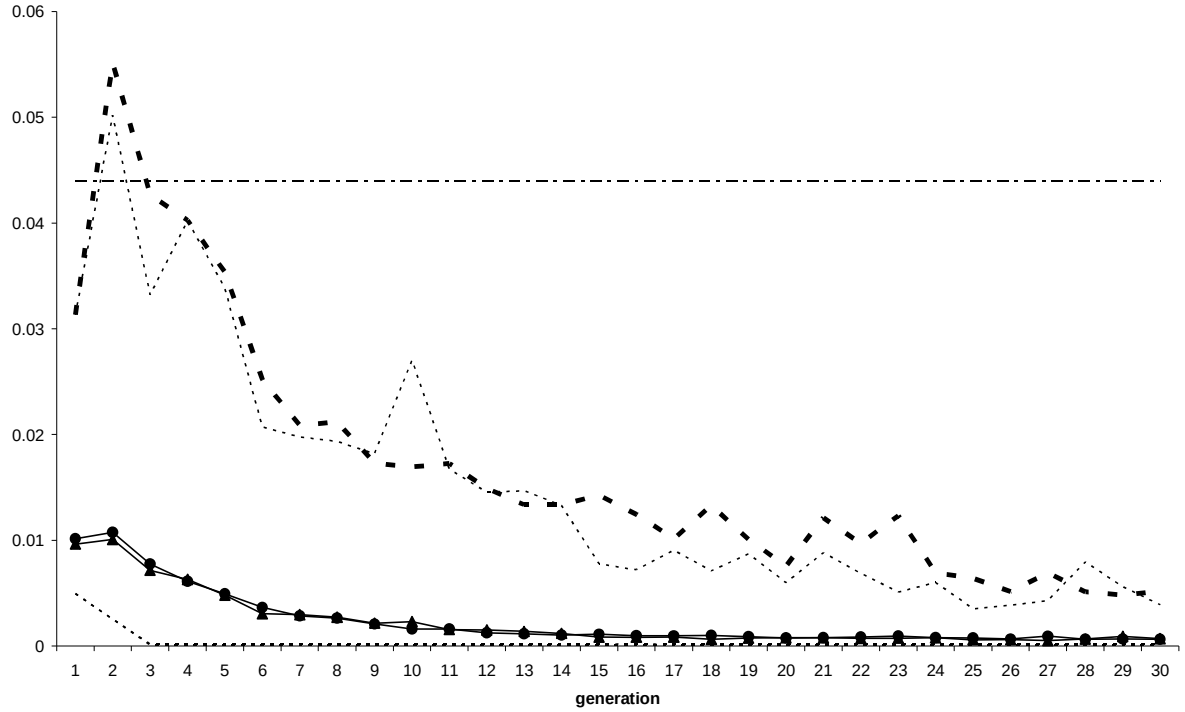


Figure 6: Risk of facing a peak over 30 generations when temporary (dotted line) or continual (plane line) counter-selection against the lethal allele occurred, assuming total number of sires to be 100, of which 10 bull-sires (triangle) or assuming total number of sires to be 25, of which 5 bull-sires (circle). Semi-dotted line refers to the estimated risk after 10 generations without selection against the lethal allele.

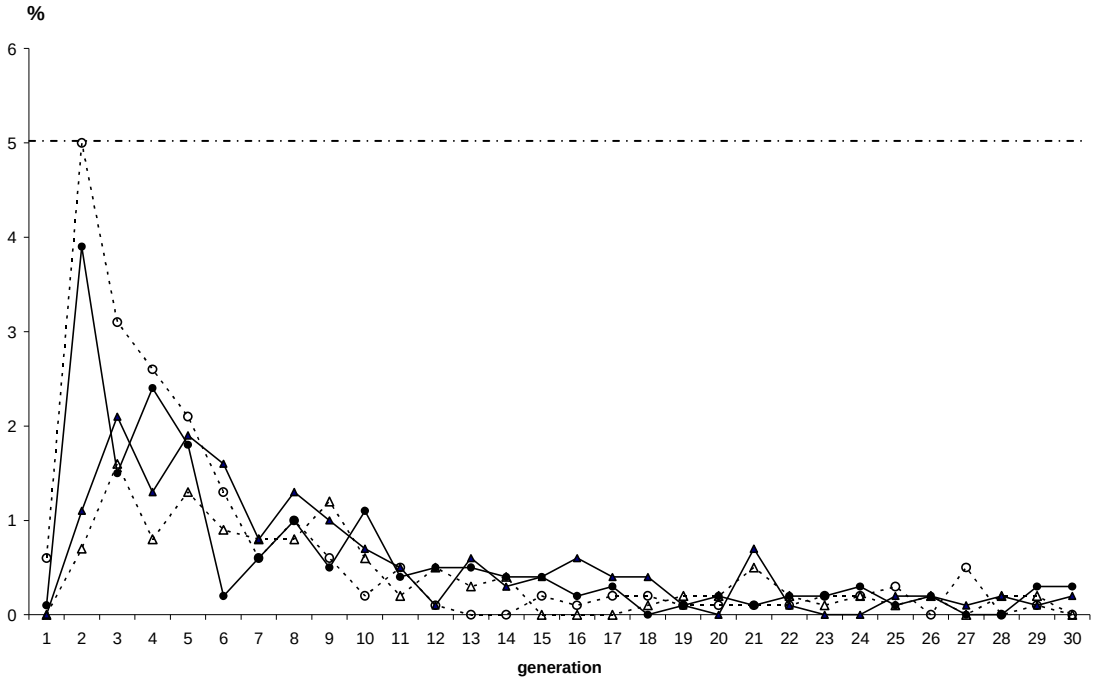
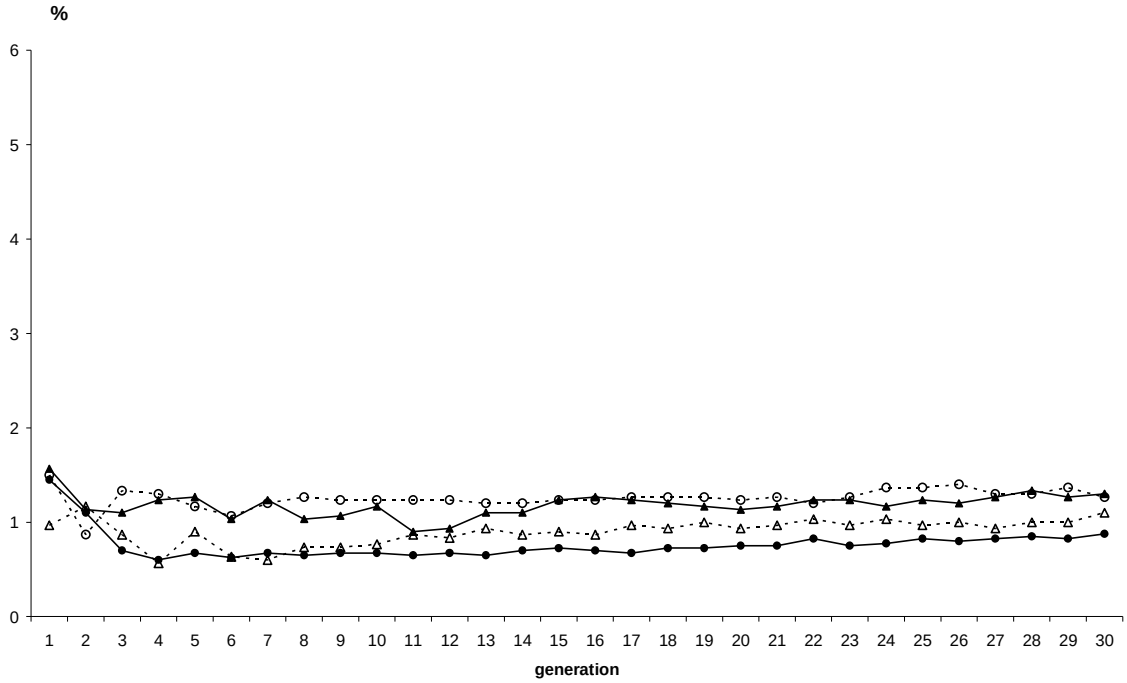


Figure 7: Risk of facing a peak at another locus over 30 generations when temporary (dotted line) or continual (plane line) counter-selection occurred against the lethal allele, assuming total number of sires to be 100, of which 10 bull-sires (triangle) or assuming total number of sires to be 25, of which 5 bull-sires (circle).



GENERAL DISCUSSION

The purpose of this thesis was to investigate in details genetic variability, using different kinds of information, such as genetic markers, especially the molecular markers, pedigrees, or phenotypes for quantitative traits, and also using different methodologies based on statistical analysis or simulations.

1. Interest of analysis of genetic variability at different scales

Investigation of genetic variability could be handled at different scales according to the types of information and the methods to analyse them. The scales are defined according to the part of the genome concerned and depend essentially on the type of data available.

The criteria of variability based on pedigree data (chap. 2) represent, in a probabilistic way, the polymorphism at an anonymous locus, strictly speaking a neutral locus with no mutation: the results may be generalised to any part of the genome following the same assumptions. The genetic parameters (chap. 3), i.e. the parameters defined within the framework of the usual model for representing the phenotypic variation of a given trait (additive variance, heritability, etc.), provide a picture of the variability due to an unknown number of unknown functional genes. The criteria of variability established for given known loci, functional (chap. 3) or not (chap. 4), give a measure of the variability directly on the DNA but restricted to the concerned loci and, to a lower extent, to the vicinity of these loci.

Within a population, the evolution of the above criteria from one generation to the other, or from one cohort of animals to the other, generally allows to quantify the evolution of the genetic variability over time. Such information are very useful for people in charge of the genetic management of the population. For this reason, parameters as the rate of inbreeding, the change in the expected heterozygosity or the temporal variation of allele frequencies, are often computed on real populations. Note, however, that the current methods used to estimate the genetic parameters (e.g., REML) always refer to the variance in the base population: using such methods in order to quantify the trend in genetic variance tells complex methodological questions.

The first interest of analysis genetic diversity at different scales lies in getting different but complementary points of view so that results may be contrasted or nuanced. In chapter 3 for instance, comparing the two hypotheses about the effect of MHC (fixed effect vs. random effect) or comparing the two methodologies about its allele frequencies, let us know if the variations for this locus may be large enough to be detected.

The second interest of analysis genetic diversity at different scales lies in focusing on a gradual reduction of the number of relevant loci. In chapter 4 for instance, a first population-scaled analysis such as PCA gives a good overview of the divergence of the lines and is also useful to find out which markers are mainly involved in the divergence; second, a genome-scaled analysis using polymorphism evolution of loci over time (*fc*) and fixation indices (*Fis*) gives a set of markers that may be influenced by selection. Finally, simulations of allele polymorphism, *i.e.* an analysis restricted to some loci, confirm which marker was actually under the influence of selection.

2. Estimation of effective size of population

The effective size (*Ne*) of the population is a key parameter for estimating the trend in genetic variability. *Ne* is usually estimated using pedigree information and the rate of inbreeding (Ne_I) or from the variance in allele frequencies over time (Ne_V); both methods were used in this study. Compared to other descriptive parameters based on pedigree information, such as effective number of founders or ancestors, rate of inbreeding based on pedigree information appeared in chapter 2 to be the most appropriate parameter for estimation of effective size of population Ne_I.

In different studies in this thesis, the estimated values of the effective size were always lower when based on inbreeding (Ne_I) than when based on the temporal variation approach (Ne_V), whatever the loci that were considered. This was true for a candidate gene whose neutrality was tested (chap. 3), and for supposedly neutral or selected markers (chap. 4). Such results let suppose that the change in genetic variability may be underestimated using pedigree information in comparison with molecular data. Moreover, estimation using genotype information from supposedly neutral marker was lower than estimation using genotype

information from markers under selection (chap. 4), meaning picture of genetic variability will depend on the nature of molecular information.

Note that estimation of the effective size population based on temporal allele frequencies at MHC locus has to be toned down because of the balancing selection that occurs in this region. Picture of genetic variability based on molecular information should however be nuanced when it deals with microsatellites because it has be shown that amplification success declined with genetic distance from the base population (Wright *et al.*, 2004), inducing a bias in the use of microsatellites to assess neutral variability. This study was undergone in flies' population but similar patterns have been found in studies about birds (Primmer *et al.*, 1996; Galburesa *et al.*, 2000). Moreover, successive generations led to drop in variability (reduction of Ne_V) that could be attributed to loss of rare alleles, rare alleles being lost before common alleles

Different temporal estimates of *Ne* were compared in mass selection lines in Pacific oysters (Appleyard and Ward, 2006): the moments temporal method that we have used along this thesis (Waples, 1989), a Bayesian approach (Bertier *et al.*, 2002) and a pseudo-likelihood approach (Wang, 2001). It appeared that the three temporal methods gave very similar estimations of *Ne*, differences being mainly due to stochastic effects associated with each model. Then, Ne_V could be calculated using any of the three methods based on microsatellites information.

Although, opposite results about genetic variability using different kinds of information were observed in quails (Kim *et al.*, 2007) and in plant breeding (Soleimani *et al.*, 2002). Estimation of inbreeding in Japanese quail lines was more accurate (smaller standard deviation) using pedigree analysis than microsatellites analysis; and in cultivars of wheat, genetic distances based on pedigree analysis (through kinship coefficient) were larger than those based on molecular information (AFLP markers – amplified restriction fragment polymorphism). But in both cases, knowledge of the pedigree was not complete or could not be traced back for more than a couple of generations, which may overestimate the actual level of variability.

3. Benefits and limits of modelling

Computer simulations are a useful tool in mimicking reality. But one question arises when modelling: what genetic model should be considered? a detailed and complex model, in order to be as realistic as possible and fit the real dataset, or a simple but unrealistic one?

Predictions of genetic gain and inbreeding (chap. 2) were calculated using deterministic equations assuming simplified hypotheses, such as polygenic additive model, non-overlapping generations, random mating. Hypothesis of drift as the only force affecting evolution of allele frequencies could be also tested (chap. 3 and 4) using a simple model: the observed parameter (*fc*) was compared with the distribution of *f* obtained from a series of simulations of a population undergoing pure drift and that presents the same initial allelic frequencies and the same inbreeding effective size. This test allowed us to identify loci with extreme *f* values compared to the rest of the genome, i.e. values greater then expected values under drift alone. In both case, modelling was an efficient approach to make useful predictions of the evolution of selected populations although the basic assumptions considered in the genetic model are not met in reality. In chapter 5, computer simulations of a simplified model that mimics the selection and reproduction scheme of dairy cattle allowed us to understand why fast increase of lethal allele frequency occurred. But consequences on genetic diversity that were drawn out of this study actually minimize the real effect of intensive artificial selection and overlapping generations and then, minimize risks.

However, more complex models may be useful and give more accurate analysis of genetic diversity. In chapter 4, for instance, the selection scheme was taken into account. Knowledge of the complete pedigree and phenotypes of the selected trait for each individual, but also the specific associations between markers, allowed us to fit reality and be able to decipher whether polymorphism evolution of the markers was under selection force or not, which could not clearly defined using statistical analysis.

4. Utility and improvement of experimental dataset

The advantages of our experimental lines were manifold:

- Pedigree was complete and we had even knowledge of founders population that were crossed to create the base population, so that we could get precise and unbiased measures of genetic variability using pedigree information, even on a long-term scale since the number of selected generations was large.
- The number of individuals and the number of parents were constant over time so that it fit basic assumptions of theoretical approaches.

- All individuals were measured for selection criteria, even for the trait they were not selected for, so that comparison between control and selected lines could be handled but also the estimation of a selected trait on other traits.
- All individuals were blood-sampled so that molecular information could (in theory, i.e. without financial constraints) be available at any generation.

Then, these experimental chicken lines have allowed us to validate and compare theoretical predictions with real data as well as to investigate in details genetic variability, using different kinds of information. However, selection rules were not strictly those of truncation selection: because of the small size of the population because, representation of the half-sib families had to be maintained and some individuals were kept in spite of their lower phenotypes.

Moreover, this experiment was not drawn on purpose for genetic diversity study: the first goal of those experimental lines was to improve selection for immune response in chickens. If we could set up a new experiment aiming at analysing genetic diversity, advice could be given:

- To combine molecular information from single nucleotide polymorphisms (SNPs) and microsatellites. SNPs are biallelic markers with a limited information content of a single marker but their high density makes them the marker of choice for linkage disequilibrium studies and then, would be useful for detection of selection signature. On the contrary, microsatellites have a low density due to the limited number of microsatellites in the chicken genome but they offer the advantage of a multiallelic marker, which is highly informative.
- To breed replicate lines so that estimation of genetic parameters would be an average over independent data from the replicates and thus, bias in estimation of genetic diversity would be reduced. However, this piece of advice may often not be taken due to resources limitations (financial, material or staff means).
- To reduce the number of generations and prefer a larger number of individuals per generation so that results would fit better with selected populations and detection of signature of selection could be handled even for low-effect QTL or low-heritability trait. In that case, another type of analysis could be undertaken for the identification of selective sweep such ln*RV*-test (Schlötterer, 2001) based on the ratio of observed variances in repeat number in two populations: this multiloci test is specially adapted to microsatellites and its power to identify a selected locus is stronger in the case of comparison of closely related populations that have faced a recent reduction in variability. This method was recently used about sunflower hybrid species (Edelist *et*

al., 2006): variability of microsatellites linked to adaptation (natural selection) QTLs appeared to be lower than those of neutral regions, as we have shown previously. However, selection experiments provide the framework for the study of the inheritance of complex traits and allow the evaluation of theoretical predictions by comparing observations against expectations. Depending on the time scale, the objectives of selection experiments may differ: short-term experiments can be used, for instance, to estimate genetic variances and covariances, and estimate the magnitude of the initial rates of response to selection. Long-term experiments are useful in monitoring changes in the rates of response or variances caused by selection itself. As these changes are dependent on the number, effects and frequencies of the genes which influence the quantitative trait, long-term experiments may provide more detailed information about its underlying inheritance. Finally, long-term selection studies may uncover evolutionary results that are different from those seen in earlier generations (Hill, 1980; Hill and Caballero, 1992).

References

- Appleyard S.A., Ward R.D. (2006). *Genetic diversity and effective population size in mass selection lines of Pacific oyster (Crassostrea gigas)*. Aquaculture **254**, 148-159.
- Avon L., Colleau J.J. (2006). *Conservation in situ de 11 races bovines françaises à très faibles effectifs : bilan génétique et perspectives*. Rencontres Recherche Ruminants **13**.
- Berthier P., Beaumont M.A., Cornuet J-M., Luikart G. (2002). *Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach*. Genetics **160**, 741-751.
- Boichard D., Maignel L., and Verrier E. (1997). *The value of using probabilities of gene origin to measure genetic variability in a population*. Genetics Selection Evolution **29**, 5-23.
- Bovenhuis H., Bralten H., Nieuwland M.G.B., Parmentier H.K. (2002). *Genetic parameters for antibody response of chickens to sheep red blood cells in a selection experiment*. Poultry science **81**, 309-315.
- Bulmer M.G. (1971). *The effect of selection on genetic variability*. Am Nat. **105**, 201-211.
- Caballero A., Toro M. (2000). *Interrelations between effective population size and pedigree tools for the management of conserved populations*. Genetical Research **75**, 331-343.
- Colleau J-J., Moureaux S., Briend M., Bechu J. (2004). *A method for the dynamic management f genetic variability in dairy cattle*. Genetics Selection Evolution **36**, 373-394.
- Colleau J-J., Moureaux S. (2006). *Gestion optimisée de la parenté et de la consanguinité dans les programmes de sélection des bovins laitiers*. INRA Productions Animales **19**, 3-14.
- Colleau J-J., Tribout T. (2006). *Gestion optimisée de la variabilité génétique dans les populations porcines sélectionnées.* INRA Productions Animales **19**, 339-346.
- Crow J.F., Kimura M.(1970). *An introduction to population genetics theory* – chap 9, 433-478. Harper & Row, New York.
- Djellali A., Vu Tien Khang J., de Rochambeau H., Verrier E. (1994). *Evaluation of the genetic conservation programmes for the sheep breeds Solognote and Mérinos précoce*. Genetics Selection Evolution **26**, 255s-265s.

- Ducos A., Ducrocq V., Eggen A., Boichard D., Malafosse A., Regaldo D., Bazin S., Gouraud X., Calvo I., Catrou O. (2004). *Implementation of a new French program for a better control of bovine genetic defects.* 55th Annual Meeting of the EAAP, Bled, Slovenia.
- Edelist C., Lexer C., Dillmann C., Sicard D., Riesberg L.H. (2006). *Microsatellite signature of ecological selection for salt tolerance in a wild sunflower hybrid species, Helianthus paradoxus*. Molecular Ecology **15**, 4623-4634.
- Estoup A., Gharbi K., San Cristobal M., Chevalet C., Haffray P., Gyuomard R (1998). *Parentage assignment using microsatellites in turbot (Scophtalmus maximus) and rainbow trout (Oncorhyncus mykiss) hatchery populations.* Journal of Fisheries and Aquatic Sciences **55**, 715-725.
- Fabuel E., Barragan C., Silio L., Rodriguez M.C., Toro M.A. (2004). *Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellites*. Heredity **93**, 104-113.
- Falconer D.S., Mackay T.F.C. (1996). *Introduction to quantitative genetics*. Longman Scientific & Technical, Essex, UK.
- FAO (2004). *Measurement of domestic animal diversity – A review of recent diversity studies*. Commission for genetic resources for food and agriculture. Working group for genetic resources for food and agriculture. 3rd session, 31 March – 2 April 2004, Rome, Italy.
- Fernandez B., Santiago E., Toro M.A., Cabalerro A. (2000). *Effect of linkage on the control of inbreeding in selection programmes.* Genetics Selection Evolution **32**, 249-264.
- Fernandez J., Villuaneva B., Pong-Wong R., Toro M.A. (2005). *Efficiency of the use of pedigree and molecular marker information in conservation programmes.* Genetics **170**, 1313-1321.
- Foulley J.L., Chevalet C. (1971). *Méthode de prise en compte de la consanguinité dans un modèle simple de simulation des performances*. Annales de Génétique et de Sélection animale **12**, 189-196.
- Foulley J.L., Ollivier L. (2006). *Estimating allelic richness and its diversity*. Livestock Science **101**, 150–158.
- Galbusera P., van Dongen S., Matthysen E. (2000). *Cross-amplification of microsatellite primers in birds.* Conservation genetics **1**, 163-168.
- Gianola D., Fernando R. L. (1986). *Bayesian methods in animal breeding theory*. Journal of Animal Science **63**, 217.

- Glemin S. (2003). *How are deleterious mutations purged? Drift versus non random mating*. Evolution **5**7, 2678-2687.
- Goldringer, I., Bataillon, T. (2004). *On the distribution of temporal variations in allele frequency: Consequences for the estimation of effective population size and the detection of loci undergoing selection*. Genetics **168**, 563-568.
- Graser H. U., Smith S. P., Tier B. (1987). *A derivative free approach for estimating variance components in animal models by restricted maximum likelihood*. Journal of Animal Science **64**, 1362.
- Groeneveld E. (1997). *VCE4 User's Guide and Reference Manual*. Neustadt, Germany.
- Hedrick P.W., Savolainen O., Karkkainen K. (1998). *Factors influencing the extent of inbreeding depression: an example from Scots pine*. Heredity **82**, 441-450.
- Henderson C.R. (1953). *Estimation of variance and covariance components*. Biometrics **9**, 226-252.
- Hill W.G. (1972). *Effective size of populations with overlapping generations*. Theoritical Population Biology **3**, 278-289.
- Hill W.G. (1980). *Design of quantitative genetic selection experiments*. in: Robertson A. Ed., Selection experiments in laboratory and domestic animals, CAB Farnham House, Slough, UK, pp. 1–13.

Hill W.G., Caballero A. (1992). *Artificial selection experiments.* Annual Review of Ecology and Systematics **23**, 287–310.

- Huby M., Griffon L., Moureaux S., De Rochambeau H., Danchin-Burge C., Verrier E. (2003). *Genetic variability of six French meat sheep breeds in relation to their genetic management*. Genetics Selection Evolution **35**, 637-655.
- Huston K. (1993). *Heritability and diagnosis of cogenital-abnormalities in food animals*. Veterinary clinics of North America – Food animal practice **9**, 1-9.
- James J.W. (1972). Computation *of genetic contributions from pedigree*. Theoritical and Applied Genetics **42**, 272-273.
- Kim S.H., Cheng K. M-T., Ritland C., Ritland K., Silversides F.G. (2007). *Inbreeding in Japanese quail estimated by pedigree and microsatellite analyses.* Journal of Heredity **98**, 378-381.
- Lacy R.C. (1989). *Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalence*. Zoo Biology **8**, 111-124.

- Lacy R.C. (1995). *Clarification of genetic terms and their use in the management of captive populations*. Zoo Biology **14**, 565-578.
- Lambert-Derkimba A., Casabianca F., Verrier E. (2006) *Des vaches, des hommes, des terroirs … et des qualification*s. Ethnozootechnie **79**, 27-38.
- Laval, G., Iannuccelli, N., Legault, C., Milan, D., Groenen, M. A. M., Giuffra, E., Andersson, L., Nissen, P. H., Jorgensen, C. B., Beeckmann, P., Geldermann, H., Foulley, J. L., Chevalet, C., Ollivier, L. (2000). *Genetic diversity of eleven European pig breeds*. Genetics Selection Evolution **32**, 187-203.
- Leroy G., Rognon X., Varlet A., Joffrin C., Verrier E. (2006). *Genetic variability in French dog breeds assessed by pedigree data*. Journal of Animal Breeding and Genetics **123**, 1-9.
- Mattalia S., Barbat A., Danchin-Burge C., Brochard M., Le Mézec P., Minéry S., Jansen G., Van Doormaal B., Verrier E. (2006). *La variabilité génétique des principales races bovines laitières françaises : quelles évolutions, quelles comparaisons internationales ?* Rencontres Recherche Ruminants **13**, 239-246.
- Meuwissen T.H.E (1996). *Maximizing the response of selection with a predefined rate of inbreeding*. Journal of Animal Science **75**, 934-940.
- Meyer K., Hill W.G. (1991). *Mixed model analysis of a selection experiment for food intake in mice*. Genetical Research **57**, 71-81.
- Moureaux S., Verrier E., Ricard A., Meriaux J.C. (1996). *Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphisms*. Genetics Selection Evolution **28**, 83-102.
- Nagahata H. (2004). Bovine *Leukocyte Adhesion Deficiency (BLAD): a review*. Journal of Veterinary Medical Science **66**, 1475-1482.
- Nei M. (1975). *Molecular population genetics and evolution*. North Holland publishing company, Amsterdam.
- Nei, M. (1978). *Estimation of average heterozygosity and genetic distance from a small number of individuals*. Genetics **89**, 583-590.
- Nei, M., Tajima, F. (1981). *Genetic drift and estimation of effective population size*. Genetics **98**, 625-640.
- Nomura T. (2005). *Methods for minimizing the loss of genetic diversity in conserved populations with overlapping generations*. Conservation Genetics **6**, 655-663.

- Palhiere I., Bibe B., San Cristobal M., Abdallah J., Bed'Hom B., Pantano T., Fregeat G. (2005). Une race, quatre rameaux "indépendants" : quelle diversité génétique après 30 ans ? 12th Rencontres Recherche Ruminants **12**, Poster.
- Pollak, E. (1983). *A new method for estimating the effective population size from allele frequency changes*. Genetics **104**, 531-548.
- Primmer C.R., Moller A.P., Ellegen H. (1996). *A wide-range survey of cross-species microsatellite amplification in birds.* Molecular Ecology **5**, 365-378.
Erratum in: Molecular Ecology (1997) **6**, 101.
- Rochambeau H. de, Chevalet C. (1985). *Minimisation des coefficients de consanguinité moyens dans les petites populations d'animaux domestiques*. Genetics Selection Evolution **17**, 459-480.
- Schlötterer C. (2002). *A microsatellite-based multilocus screen for the identification of local selective sweep*. Genetics **160**, 753-763.
- Searle S.R. (1989). *Variance components – some history and a summary account of estimation methods.* Journal of Animal Breeding and Genetics **106**, 1-29.
- Slatkin M. (1995). *A measure of population subdivision based on microsatellite allele frequency*. Genetics **139**, 457-462.
- Soleimani V.D., Baum B.R., Johnson D.A. (2002). *AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat [Triticum turgidumL. Subsp. Durum (Desf.) Husn]*. Theoritical and Apllied Genetics **104**, 350-357.
- Sorensen D.A., Kennedy B.W. (1984). *Estimation of genetic variances from unselected and selected populations*. Journal of Animal Science **59**, 1213-1223.
- Thaon d'Arnoldi C., Ollivier L., Foulley J.L. (1998). *An overview of the Weitzman approach to diversity*. Genetics Selection Evolution **30**, 149-161.
- van der Werf J.H.J., de Boer I.J.M (1990). *Estimation of additive genetic variance when base populations are selected*. Journal of Animal Science **68**, 3124-3132.
- van der Werf J.H.J. (1992). *Restricted maximum likelihood estimation of additive genetic variance when selected base animals are considered fixed*. Journal of Animal Science **70**, 70-1088.
- van Raden P.M. (1992). *Accounting for inbreeding and crossbreeding in genetic evaluation for large populations*. Journal of Dairy Science **75**, 3136-3144.
- Verrier E., Colleau J.J., Foulley J.L. (1989) *Effect of mass selection on the within-family genetic variance*. Theoritical and Applied Genetics **77**, 142-148.

- Verrier E., Colleau J-J., and Fouley J-L. (1994). *Genetic variability when selecting on the animal model BLUP*. 5th World Congress on Genetics Applied to Livestock Production, August 7-12, 1994, Guelph, Canada **19**, 139-142.
- Verrier E., Moureaux S., Boichard D., Danchin-Burge C., Avon L. (2001). *Gérer la variabilité génétique des populations d'élevage : l'exemple des races bovines françaises, depuis les races en conservation jusqu'aux races nationales et internationales*. 6ème Carrefour des Productions Animales, Gembloux, 24 janvier 2001, 43-51.
- Verrier E., Tixier-Boichard M., Bernigaud R., Naves M. (2005). *Conservation and values of local livestock breeds: usefulness of niche products and/or adaptation to specific environments*. Animal Genetic Resources Information **36**, 21-31.
- Verrier E., Moureaux S., Tribout T., Delaunay I., Palhière I., Rochambeau H. de, Colleau J.J. (2005). *Overview of the genetic variability in French selected livestock populations and management approaches*. Options and strategies for the conservation of farm animal genetic resources, Agropolis international workshop, CD-Rom, 28-33.
- Wahlund, S. (1928). *Zuzammensetzung yon Population und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus Betrachtet*. Hereditas **11**, 65-105.
- Wang J. (2001). *A pseudo-likelihood method for estimating effective population size from temporally spaced samples*. Genetical Research **78**, 243-257.
- Waples, R.S. (1989). *A generalized approach for estimating effective population size from temporal changes in allele frequency*. Genetics **121**, 379-391.
- Weir B.S., Cockerham C.C. (1984). *Estimating F-statistics for the analysis of population structure*. Evolution **38**, 1358-1370.
- Weitzman M.L. (1992). *On diversity*. Quarterly Journal of Economics **107**, 363-405.
- Wright S. (1921). *Systems of mating*. Genetics **6**, 111-178.
- Wright S. (1969). *Evolution and the genetics of populations, The Theory of Gene Frequencies* **vol. 2**, University of Chicago Press, Chicago, Illinois, USA.
- Wright T.F., Johns P.M., Walters J.R., Lerner A.P., Swallow J.G., Wilkinson G.S. (2004). *Microsatellite variation among divergent populations of stalk-eyed flies, genus Cyrtodiopsis*. Genetical Research **84**, 27-40.