



**HAL**  
open science

# Comparaison de méthodes d'évaluation de la diversité génétique chez un ensemble de populations d'une même espèce (*Sus Scrofa*)

Natacha N. Nikolic

► **To cite this version:**

Natacha N. Nikolic. Comparaison de méthodes d'évaluation de la diversité génétique chez un ensemble de populations d'une même espèce (*Sus Scrofa*). Sciences du Vivant [q-bio]. 2005. hal-02831278

**HAL Id: hal-02831278**

**<https://hal.inrae.fr/hal-02831278>**

Submitted on 7 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PAUL SABATIER

Master 2 Recherche  
Biodiversité, Ecologie et Evolution

Soutenue publiquement le 28 juin 2005 devant la commission d'examen.

**Natacha Nikolic**

**Comparaison de méthodes  
d'évaluation de la diversité génétique  
chez un ensemble de populations  
d'une même espèce  
(*Sus scrofa*)**



*Tuteurs de stage* : Claude Chevalet (LGC)  
Magali SanCristobal (LGC)

*Rapporteurs* : Lounès Chikhi (EDB)  
Mark Hewison (CEFS)



INRA  
Laboratoire de Génétique Cellulaire  
Chemin de Borde Rouge  
31326 Castanet Tolosan cedex



# Sommaire

<b>Résumé - Abstract</b>	<b>3</b>
<b>Remerciements</b>	<b>4</b>

---

## Volume 1 : Texte principal

---

<b>I Introduction</b>	<b>5</b>
<b>II Matériels et méthodes</b>	<b>7</b>
2.1 Méthodes en génétique conservative	
2.1.1 Méthode de Caballero et Toro (2002)	
2.1.2 Méthode d'Eding et Meuwissen (2001 - 2002)	
2.1.3 Méthode de Weitzman (1992 et 1993)	
2.2. Simulations	
2.2.1 Configurations des simulations	
2.2.2 Précisions supplémentaires	
2.3 Analyse des populations porcines	
<b>III Résultats</b>	<b>18</b>
3.1 Résolutions mathématiques	
3.1.1 Similarité entre la méthode de Caballero et Toro et celle d'Eding et Meuwissen	
3.1.2 Lien indirect entre les trois méthodes	
3.1.3 Schéma en étoile	
3.2 Simulations	
3.2.1 Comparaison de la méthode de Caballero et Toro et celle de Weitzman	
3.2.1.1 Configuration avec des populations sous dérive génétique	
3.2.1.2 Configuration avec l'ensemble des populations à l'équilibre	
3.2.2 Effet du nombre de populations	
3.2.2.1 Populations sous dérive génétique	
3.2.2.2 Populations à l'équilibre dérive et mutation	
3.2.3 Effet du coefficient de parenté	
3.3 Analyse des populations porcines	
<b>IV Discussion</b>	<b>38</b>
<b>V Conclusion</b>	<b>41</b>
<b>VI Références</b>	<b>42</b>

---

## Volume 2 : Annexes

---

<b>Annexes calculs</b>	<b>45</b>
<b>Annexes estimations</b>	<b>49</b>

## Résumé

Nos travaux se placent dans le contexte de la génétique de la conservation avec pour cas d'étude une espèce d'élevage (*Sus scrofa*). Pour cela, nous avons évalué les performances des outils couramment utilisés pour maximiser la diversité génétique d'une espèce: la méthode de Caballero et Toro (2002), d'Eding et Meuwissen (2001), et de Weitzman (1992). Les simulations et les calculs théoriques ont permis d'étudier le comportement de ces trois méthodes selon que les populations sont à l'équilibre entre dérive et mutation ou sont seulement soumises à la dérive génétique. Les résultats mettent en évidence l'importance des taux d'homozygoties et des liens de parentés entre les populations. Parmi les trois méthodes étudiées, la méthode d'Eding et Meuwissen est équivalente en espérance à la méthode de Caballero et Toro. Concernant la méthode de Weitzman, dans la plupart des cas d'études, elle évalue des contributions allant à l'encontre des deux autres méthodes : l'estimation des contributions n'est pas proportionnelle aux taux d'hétérozygotie. Ainsi, les populations consanguines sont conservées en dépit de leur faible taux d'hétérozygotie et richesse allélique. Néanmoins dans le cas de population de grande taille à l'équilibre dérive et mutation, la méthode de Weitzman se comporte comme la méthode de Caballero et Toro. L'attribution des contributions est proportionnelle au taux d'hétérozygotie. Pour finir, les trois méthodes ont été appliquées sur soixante populations porcines européennes et les analyses révèlent des résultats qui ne sont ni homogènes, ni cohérents entre les trois méthodes.

Ces travaux font ainsi émerger la problématique du choix de l'outil analytique en fonction de la structure et de la dynamique des populations. Des améliorations semblent encore nécessaires pour construire, en toute sécurité, un projet de conservation s'appuyant sur des méthodes de maximisation de la diversité génétique d'une espèce.

## Abstract

Our works take place in the context of conservation genetics in the case of one livestock specie (*Sus scrofa*). The goal of this study was to evaluate the performance of some methods maximizing genetic diversity: the method of Caballero and Toro (2002), Eding and Meuwissen (2001), and Weitzman (1972). The behaviour of methods was studied using real data (sixteen European pig populations), simulations and analytical approach following populations in mutation-drift equilibrium or under drift only. The results show the importance of homozygosity rate and of the genetic relationship between populations to determine the contributions of each population to overall diversity. The method of Eding and Meuwissen is equivalent to the method of Caballero and Toro. In most cases, the method of Weitzman gives contributions which are opposite to those given by the other two methods: the estimation of contributions is not proportional to the heterozygosity rate. The population will be conserved despite of low heterozygosity and allelic richness. In the case of large population in mutation-drift equilibrium, the method of Weitzman behaves as the method of Caballero and Toro. The contributions are proportional to the heterozygosity rate. Although they seem satisfactory in theory and in some simulations cases, the methods do not give entirely satisfactory results when applied to the sixty European pig populations. This work highlights the question of choosing the right analytical tools in function of the structure and the dynamic of populations. Improvements seem still necessary to construct conservation project aimed at maximizing genetic diversity of species.

# Remerciements

Je veux tout d'abord remercier les personnes sans qui ce travail n'aurait pas vu le jour, mes responsables de stage, Monsieur Claude Chevalet et Madame Magali SanCristobal, pour l'encadrement scientifique de très grande qualité qu'ils ont su me donner.

Je voudrais également les remercier pour leur soutien et leur aide précieuse en programmation, en analyse des données et en mathématiques dont quelques concepts m'avaient échappé. Leur lecture de ce présent mémoire m'a permis de soumettre des points supplémentaires cruciaux lors de ma discussion. Qu'ils trouvent en ces quelques lignes le témoignage de ma sincère gratitude.

Je remercie le programme Européen PigBioDiv pour avoir mis à ma disposition les données génétiques des soixante populations porcines étudiées au cours de ces travaux.

Je veux aussi adresser de vifs remerciements à Madame Solange Casette pour les arrangements de prêts d'ouvrage entre bibliothèques qui nécessite parfois des négociations très prenantes.

Je remercie Monsieur Eddie Iannuccelli et Monsieur Patrice Dehais pour leurs arrangements et réarrangements informatiques.

Mes remerciements s'adressent également à Monsieur Sovan Lek et Monsieur Young Park pour leur aide et leur soutien en matière de réseaux de neurones.

J'adresse ma sincère gratitude aux professeurs et aux intervenants du Master Recherche pour leurs conseils et le temps qu'ils nous ont consacré.

Enfin, les derniers mais les premiers dans mon cœur, je remercie ma famille pour leur indispensable soutien, mais surtout pour avoir pris avec bon côté ma *bonne humeur* lorsque j'ai rédigé ce mémoire.

## I Introduction

La biodiversité, ou diversité biologique, est l'un des thèmes grandissant dans le domaine des sciences. Elle fait référence à plusieurs caractéristiques d'un système incluant le nombre d'espèces, la quantité de variation génétique, et la complexité des interactions (Krohne, 2001). La biodiversité acquiert des définitions diverses et variées en fonction de l'échelle à laquelle on se place (Townsend *et al.* 2001). Les diverses définitions à son égard ainsi que l'absence de valeurs formelles pour la quantifier, en fait un thème complexe (Euler, 1999).

Durant nos travaux, nous nous sommes intéressés à une des composantes de la biodiversité, la diversité génétique appréhendée au moyen de marqueurs génétiques supposés neutres (séquences microsatellites). Les stratégies de conservation génétique cherchent à maximiser cette diversité génétique pour diminuer le risque d'extinction des espèces ou des populations (Nunney, 2000). La réduction de la diversité génétique entraînerait à court terme une diminution de la valeur adaptative (fitness) et à long terme une diminution du potentiel évolutif (capacité à s'adapter aux changements environnementaux). La diversité génétique peut être notamment réduite par des phénomènes de goulot d'étranglement, d'effet fondateur, d'endogamie et de dérive génétique. Au cours de ces travaux, nous nous sommes efforcés de comprendre le comportement de trois méthodes, couramment utilisées dans les projets de conservation des animaux d'élevage. Nous avons analysé leur comportement selon que les populations sont sous dérive génétique ou à l'équilibre entre dérive et mutations. S'interroger sur les limites de certaines méthodes d'évaluation de la variabilité génétique permet une utilisation plus efficace de celles-ci.

La plupart des populations porcines européennes subissent de la dérive génétique (Nikolic *et al.* 2009 ; SanCristobal *et al.* 2005) et sont ainsi des modèles adaptés pour étudier cet effet. La dérive génétique est un déplacement erratique des fréquences alléliques, dû à l'effet du hasard (Binder, 1967 ; Nei, 1987). De génération en génération, la fréquence d'un gène varie et il peut arriver que cette fréquence atteigne les valeurs, 0 ou 1, (élimination ou fixation du gène au sein de la population). Il existe donc des écarts de fréquence d'une génération à l'autre. La variance faisant intervenir la taille de la population, cet écart est d'autant plus important que l'effectif de la population est faible (Binder, 1967). La dérive génétique est donc beaucoup plus sensible lorsque les populations sont petites, entraînant une probabilité de fixation ou d'élimination de gènes plus élevée (Binder, 1967). Afin de mesurer

l'influence de la dérive génétique sur les trois méthodes d'études, nous avons fait varier la taille des populations au cours de nos simulations. Les trois méthodes analysés concernent la méthode de Caballero et Toro (2002) ; d'Eding et Meuwissen (2001), et de Weitzman (1992-1993). Les deux premières méthodes évaluent la diversité génétique par le taux d'hétérozygotie, qui est la mesure classique de quantification de la diversité génétique. La troisième se base sur des distances génétiques.

Il existe cependant une autre quantification, tout autant utilisée, fondée sur l'évaluation du nombre d'allèle(s) différent(s) présent(s) dans une population (la richesse allélique). A l'heure actuelle, aucune étude n'a développé des stratégies de conservation pour maintenir ou maximiser la richesse allélique. Par contre, les études sur le taux d'hétérozygotie offrent des stratégies de conservation pour maintenir ce taux moyen dans les populations à un niveau élevé. Les approches présentées dans cet écrit suggèrent une maximisation de la diversité génétique par optimisation des contributions populationnelles d'une même espèce.

Pour conclure, ce mémoire s'articule en trois volets. Après un bref aperçut de chacune des méthodes étudiées au sein de ces travaux, le premier volet propose leurs analyses par des calculs théoriques. Le second volet examine le comportement de deux d'entre-elles au travers de simulations sous le modèle de mutation à nombre infini d'allèles (IAM : Infinite Allele Model). Pour finir sur la comparaison des trois méthodes lors de l'analyse de soixante populations porcines européennes génotypées avec 50 marqueurs génétiques. Nous verrons par ces trois volets, exposant les différences de résultats obtenus, que le choix de l'outil analytique est déterminant dans les décisions de conservation pour l'espèce.

## II Matériels et méthodes

### 2.1 Méthodes en génétique de la conservation

Bien que les deux méthodes, Caballero et Toro (2002) (Caballero-Toro) et Eding et Meuwissen (2001) (Eding-Meuwissen), n'aient pas utilisé les mêmes notations, nous introduisons des notations communes. Un  $f$  représentant les valeurs observées et  $F$  les valeurs théoriques.

#### 2.1.1 Méthode de Caballero et Toro (2002)

L'approche de Caballero-Toro est une approche évaluant une fonction de diversité ( $GD$ ) en passant par des moyennes arithmétiques et un algorithme d'optimisation sur la base d'une équation quadratique ( $c' \tilde{f} c$ ) (voir *Annexe A*).

$$GD = 1 - c' \tilde{f} c = 1 - \sum_{i,j=1}^n f_{i,j} c_i c_j \quad (1.1)$$

$$GD = 1 - \left( \sum_i c_i^2 f_{i,i} + \sum_{j \neq i} c_i c_j f_{i,j} \right) \quad (1.2)$$

Avec  $\tilde{f}$  matrice symétrique d'éléments  $f_{i,j}$ ,  $c_i$  contribution de la population  $i$  et  $c_j$  contribution de la population  $j$ . Une contribution peut être définie comme la proportion de la population devant être prélevée ou conservée pour maximiser la diversité génétique ( $GD$ ). Maximiser  $GD$  revient à minimiser l'équation quadratique sous les contraintes  $0 \leq c_i \leq 1$  et  $\sum_i c_i = 1$ . Les contributions sont évaluées numériquement à l'aide d'un algorithme convexe en fonction de la matrice  $f_{i,j}$  qui est calculée préalablement à partir des fréquences alléliques ( $p_{k,l,i}$ ) de chaque individu par population. La fréquence prendra la valeur 1 lorsque l'individu  $l$ , de la population  $i$ , est homozygote pour l'allèle  $k$ . Elle aura la valeur 0,5 si l'individu est hétérozygote pour l'allèle  $k$  et une valeur nulle si l'individu ne porte pas cet allèle. Ces fréquences individuelles sont ensuite sommées par population (avec  $N_i$  le nombre d'individu total dans la population  $i$ ) pour obtenir une fréquence génotypique moyenne ( $\hat{p}_{k,i}$ ):

$$\hat{p}_{k,i} = \frac{\sum_{l=1}^{N_i} p_{k,l,i}}{N_i} . \quad (1.3)$$



Ces fréquences populationnelles moyennes et le nombre d'allèles  $a$  permettent de calculer la matrice  $f_{i,j}$  :

$$f_{i,j} = \sum_{k=1}^a \widehat{P}_{k,i} \widehat{P}_{k,j} . \quad (1.4)$$

Ce n'est que lorsque la matrice symétrique est calculée que l'algorithme d'optimisation évalue les contributions en pourcentage pour chaque population. Ces contributions ont pour but de maximiser la diversité génétique en reconstituant une nouvelle population à partir des populations déjà existantes.

### 2.1.2 Méthode d'Eding et Meuwissen (2001 - 2002)

Cette approche a le même objectif que celle de Caballero-Toro en termes de maximisation de la diversité génétique. Elle fournit des contributions par population qui serviront de référence pour constituer une nouvelle population cherchant à représenter une population fondatrice commune aux populations observées.

Cependant le calcul des contributions est analytique et passe par un modèle de régression linéaire logarithmique. Les auteurs définissent l'espérance de la fonction de diversité (1.5), entre les populations  $i$  et  $j$  descendantes de la population fondatrice  $o$ . Comme nous l'expliquons en 3.1.1, ils écrivent

$$E(1 - f_{i,j}) = (1 - F_{i,j})(1 - \sum_k p_{o,k}^2) , \quad (1.5)$$

où  $F_{i,j}$  est le coefficient moyen de parenté entre individus de ( $i$ ) et de ( $j$ ),  $p_{o,k}$  la fréquence de l'allèle  $k$  de la population  $o$  et  $f_{i,j}$  a la même valeur que dans l'approche de Caballero-Toro (Equation 1.4). Puis linéarisent cette fonction (1.6 et 1.7)

$$\ln(E(1 - f_{i,j})) = \ln(1 - F_{i,j}) + \ln(1 - \sum_k p_{o,k}^2) . \quad (1.6)$$

L'équation (1.6) peut s'écrire sous la forme (1.7) où  $\mathbf{a}$  et  $\mathbf{b}$  sont des vecteurs du modèle et  $\mathbf{e}$  le vecteur de résidus qui suit une loi centrée de variance  $\sigma^2$ .

$$y_{i,j,o} = a_{i,j} + b_o + e_{i,j,o} \quad (1.7)$$

Cette approche prend en compte le taux d'homozygotie du locus considéré dans la population fondatrice ( $p_{o,k}^2$ ) pour calculer une matrice  $\hat{F}_{i,j}$  par une transformée du vecteur  $a_{i,j}$  après ajustement du modèle (1.7) :

$$\hat{F}_{i,j} = 1 - e^{a_{i,j}} . \quad (1.8)$$

Grâce à cette matrice  $\hat{F}_{i,j}$ , le vecteur des contributions optimales de  $n$  (nombre de populations) dimensions ( $C \min$ ) est évalué par l'équation suivante :

$$C \min = \frac{\hat{F}_{i,j}^{-1} \mathbf{1}_n}{\mathbf{1}'_n \hat{F}_{i,j}^{-1} \mathbf{1}_n} , \quad (1.9)$$

où  $\mathbf{1}_n$  est le vecteur de longueur  $n$  composé de 1.

Ce vecteur de contributions minimise la fonction suivante :

$$C' \hat{F} C . \quad (1.10)$$

Cette méthode est divisée en trois sous-méthodes : ULM (Unweighted Log-linear model), WLM (Weighted Log-linear Model) et WLMM (Weighted Log-linear Mixed Model) (Eding et Meuwissen, 2002).

Dans la sous-méthode ULM, les résidus sont supposés de même variance. Eding et Meuwissen ont démontré que cette hypothèse n'était pas fondée. Il est donc préférable d'utiliser les deux autres sous-méthodes.

Les sous-méthodes WLM et WLMM considèrent l'équation (1.7) avec des résidus ayant chacun une variance différente et  $a_{i,j}$  fixe pour WLM et aléatoire pour WLMM.

Au cours de nos simulations, la méthode d'Eding-Meuwissen ne sera pas utilisée. Ce choix provient de la similarité observée au cours de nos résolutions mathématiques avec l'approche de Caballero-Toro. Un Tableau récapitulatif (**Tableau 1**) rassemble les deux premières approches.

Méthodes d'optimisation  Estimation de la matrice	Numérique	Analytique
$F_{i,j} = \sum_{k=1}^a \widehat{P}_{k,i} \widehat{P}_{k,j}$ On somme sur les locus	Méthode de Caballero-Toro	Néant
ULM (Unweighted Log-linear Model)	Néant	Sous-méthode de la méthode d'Eding-Meuwissen
WLM (Weighted Log-linear Model)	Néant	Sous-méthode de la méthode d'Eding-Meuwissen
WLMM (Weighted Log-linear Mixed Model)	Néant	Sous-méthode de la méthode d'Eding-Meuwissen

**Tableau 1.** Tableau récapitulatif de la méthode de Caballero et Toro (2002) et des trois sous-méthodes d'Eding et Meuwissen (2001).

### 2.1.3 Méthode de Weitzman (1992 et 1993)

Martin Weitzman (1992) introduit le concept de fonction de diversité  $V(S)$  qui est une fonction se basant sur les différences entre populations appartenant à un même ensemble. Il suggère (1993) l'utilisation de l'information génétique pour déterminer les différences entre les populations. Pour calculer la fonction de diversité  $V(S)$ , il faut avoir au préalable évalué une distance génétique  $d(i, j)$  entre les populations  $i$  et  $j$ .

$V(S/i)$  représente la fonction de diversité sans la population  $i$ . Les contributions ( $c_i$ ) s'évaluent selon la fonction suivante de  $V(S)$  et  $V(S/i)$  :

$$c_i = 1 - V(S/i)/V(S). \quad (1.11)$$

En *Annexe B*, on pourra trouver un exemple de calcul de ces contributions.

La distance génétique entre deux populations  $d(i, j)$  est généralement donnée par la distance de Nei standard (1972) ou la distance de Reynolds (1983). Lorsque l'analyse génétique se déroule par le biais de marqueurs microsatellites et que les populations sont sous dérive génétique, la distance de Nei standard et la distance de Reynolds sont corrélées (Laval, 2002). Cependant, nous verrons que lorsque les populations sont à l'équilibre dérive et mutation, il est préférable d'utiliser la distance de Reynolds.

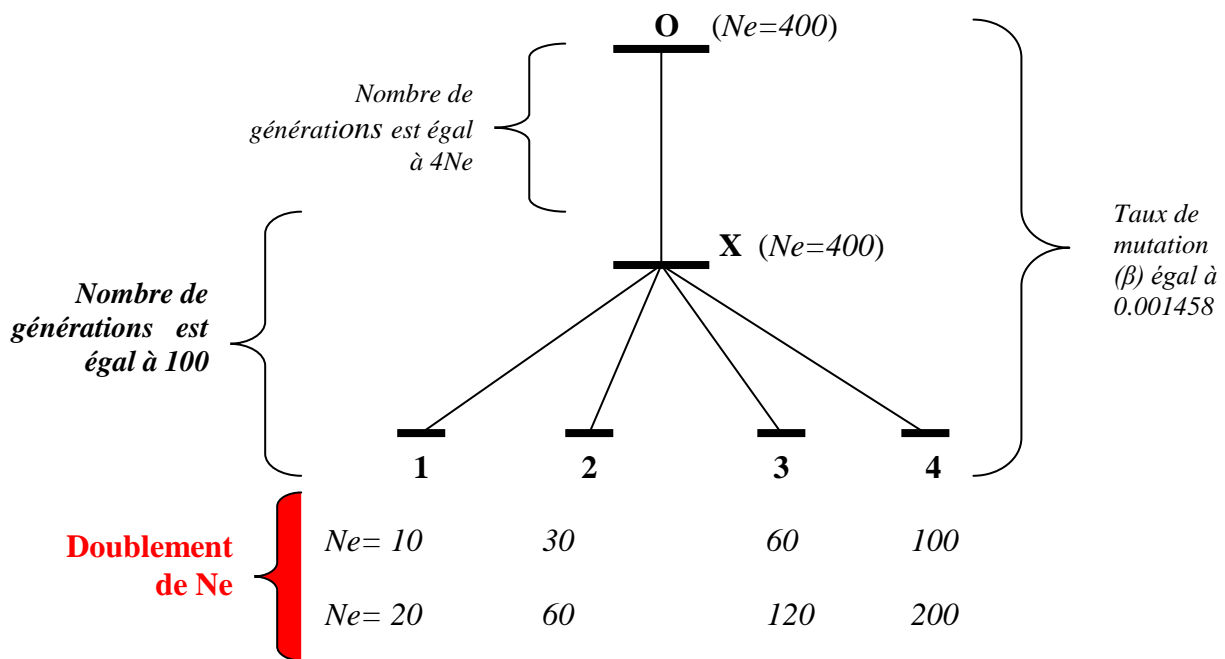
Cette méthode est mise en oeuvre par un programme informatique développé Serguei Derban (ENSAE, Paris) en collaboration avec Jean-Louy Foulley et Louis Ollivier (2002).

## 2.2. Simulations

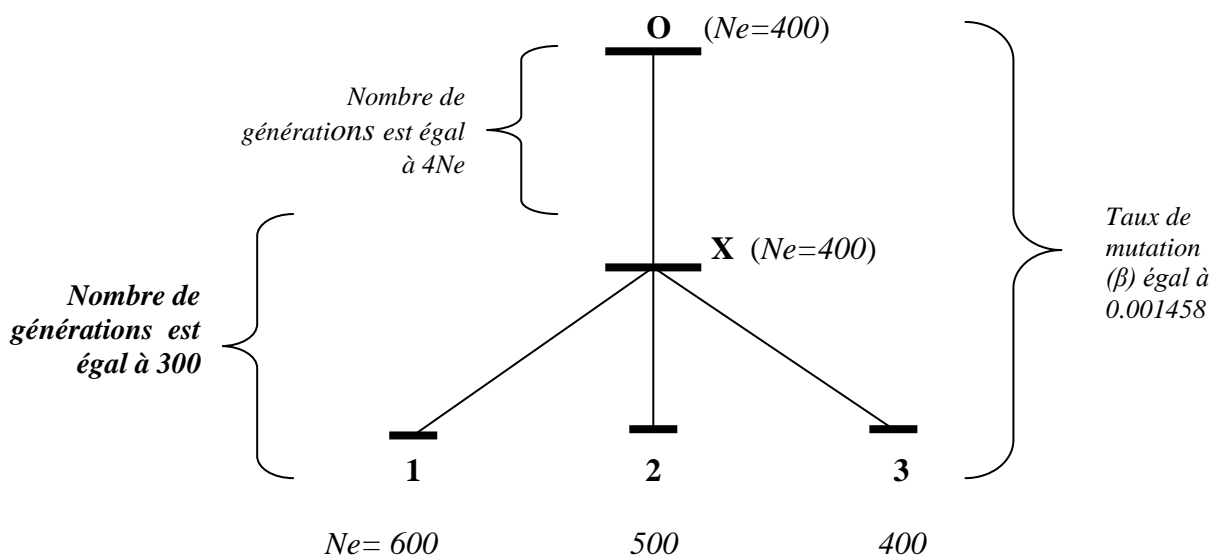
### 2.2.1 Configurations des simulations

Le travail de simulation s'est organisé en trois parties.

La première partie repose sur la compréhension du comportement de la méthode de Caballero-Toro et de la méthode de Weitzman, selon que les populations sont sous dérive génétique (*Figure 1*) ou à l'équilibre (*Figure 2*). Dans la première configuration, nous avons construit quatre populations sous dérive génétique avec des tailles efficaces différentes pour avoir des taux d'hétérozygotie différents.



**Figure 1.** Configurations avec un doublement des populations sous dérive génétique pour obtenir trois populations sous dérive génétique et une à l'équilibre dérive et mutation. Les populations descendent toutes d'une même population à l'équilibre.



**Figure 2.** Configuration de trois populations à l'équilibre dérive et mutation provenant d'une même population fondatrice à l'équilibre. La taille efficace des trois populations est différente allant de manière décroissante de la population (1) à la population (3).

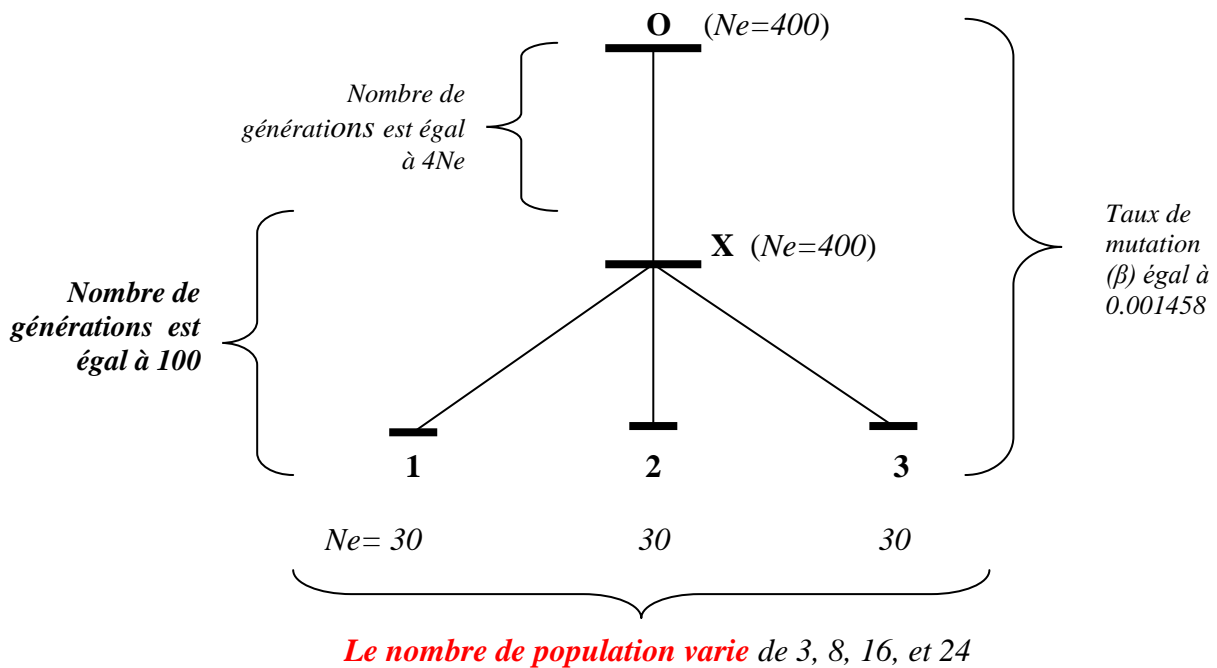
- La taille efficace de ces populations a été doublée pour étudier le comportement des approches quand on passe de quatre populations sous dérive génétique à trois sous dérive et une à l'équilibre (**Figure 1**). Dans la seconde configuration, nous avons construit trois populations à l'équilibre avec des tailles effectives différentes pour avoir des taux d'hétérozygotie décroissants des populations (**Figure 2**). Les taux d'hétérozygoties moyen des populations ont été superposés aux contributions moyennes des simulations analysées par les deux approches.

- La seconde partie s'est intéressée à l'influence du nombre de populations sur les méthodes. Pour cela, nous avons construit une configuration avec des populations sous dérive génétique (**Figure 3**) et une autre configuration avec des populations à l'équilibre (**Figure 4**). Nous avons fait varier le nombre de populations de 3 à 24 pour les populations sous dérive et de 8 à 24 pour les populations à l'équilibre construisant 1000 simulations à chaque variation du nombre de populations. Ces simulations ont été ensuite analysées par les deux méthodes, Caballero-Toro et Weitzman.

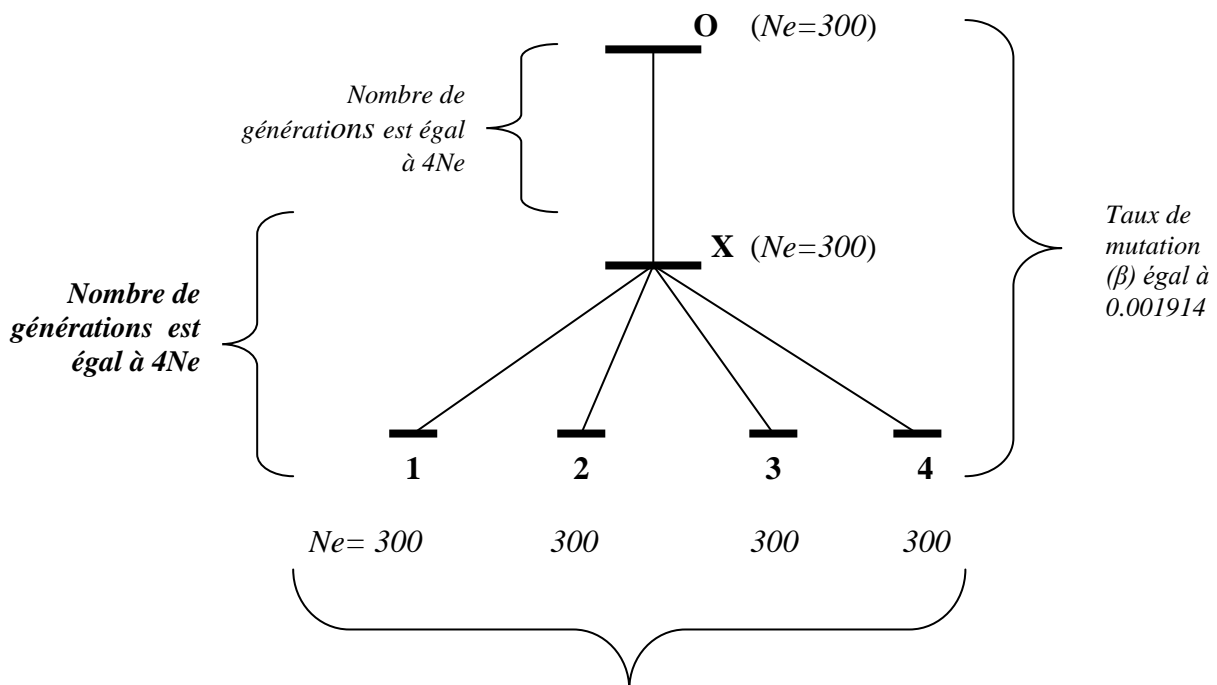
L'intervalle de génération chez le porc est d'environ un an. De plus, la sélection intensive des populations porcines par l'homme serait évaluée à un siècle (SanCristobal *et al.* 2005). Nous avons donc choisi un temps de génération égal à 100 pour les simulations plaçant les populations sous dérive génétique.

- La dernière partie avait pour objectif d'évaluer l'effet du coefficient de parenté. Pour cela, deux configurations ont été construites à partir d'une population à l'équilibre se scindant en deux. Une population (1) sur une branche et les deux autres (2 et 3) sur l'autre branche avec des temps de génération au niveau de la seconde ramification qui diffèrent. Ce temps est égal à 5 générations pour le premier modèle (**Figure 5**), entraînant une grande proximité génétique entre les populations 2 et 3 (coefficient de parenté élevé). Un nombre de générations de 200 dans la seconde configuration a été choisi (**Figure 6**) afin que les populations 2 et 3 soient génétiquement éloignées (coefficient de parenté faible). Les contributions obtenues par les deux méthodes ont été comparées avec les taux d'hétérozygotie, les richesses alléliques et les coefficients de parenté.

Ces trois parties permettent une analyse ciblée du comportement de la méthode de Caballero-Toro et de la méthode de Weitzman en fonction de la parenté entre les populations et du taux d'homozygotie des populations.

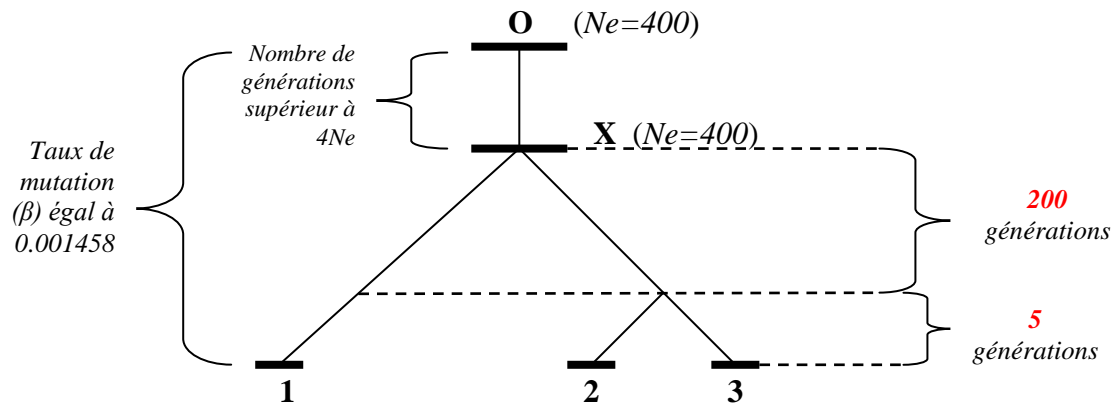


**Figure 3.** Configurations de populations sous dérive génétique ayant une taille efficace réduite ( $Ne=30$ ) et provenant d'une même population fondatrice à l'équilibre dérive et mutation. Quatre configurations ont été construites de sorte de faire varier le nombre de populations descendantes de 3 à 24 populations.

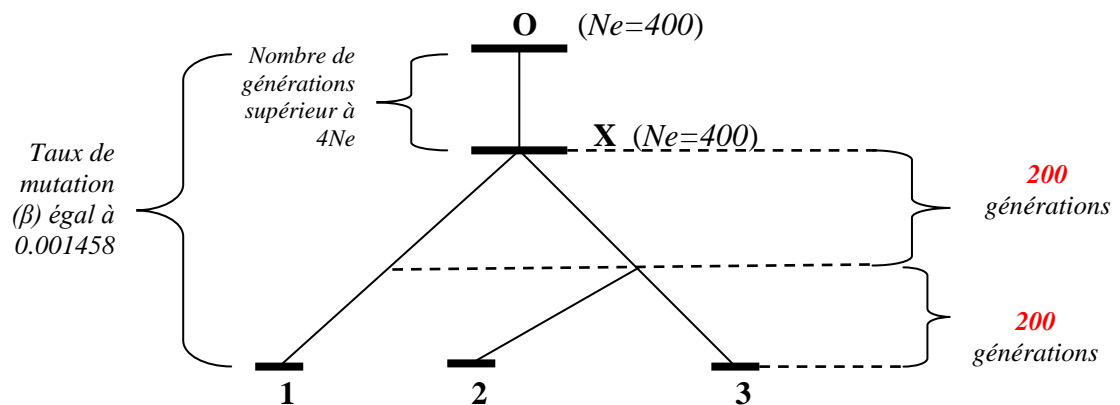


**Figure 4.** Configurations de populations à l'équilibre dérive et mutation ayant une taille efficace élevée ( $Ne=300$ ) et provenant d'une même population fondatrice à l'équilibre dérive et mutation. Quatre

configurations ont été construites de sorte de faire varier le nombre de populations descendantes de 8 à 24 populations.



**Figure 5.** Modèle de trois populations provenant d’une population fondatrice à l’équilibre. La population originelle s’est scindée en deux clades dont l’un s’est ramifié en deux populations (2 et 3) génétiquement proches. La taille efficace des trois populations est la même ( $Ne=30$ ) et le taux de mutation est fixé.



**Figure 6.** Modèle de trois populations provenant d’une population fondatrice à l’équilibre. La population originelle s’est scindée en deux clades dont l’un s’est ramifié en deux populations (2 et 3) génétiquement proches. La taille efficace des trois populations est la même ( $Ne=30$ ) et le taux de mutation est fixé.



### 2.2.2 Précisions supplémentaires

L'ensemble des simulations a été réalisé par un programme permettant de décrire la transmission des allèles au sein d'un groupe de populations dont la structure est spécifiée par l'utilisateur (Chevalet et Rochambeau, 1985 ; programme adapté pour intégrer les phénomènes de dérive et mutation, selon le modèle IAM (Infinite Allele Model)). Ce modèle introduit par Malécot (1946), puis largement utilisé par Kimura et Crow (1964), et Nei (1971, 1972, 1975), néglige les mutations reverses (probabilité qu'un allèle mute vers un état préexistant dans la population). Toute nouvelle mutation donne lieu à un nouvel allèle. Les allèles identiques proviennent forcément d'un même allèle ancestral et ils sont dits identiques par descendance (IBD : Identical By Descent). Le modèle IAM en comparaison avec le modèle de mutation pas à pas (SMM : Stepwise Mutation Model, qui serait mieux adapté pour représenter les marqueurs microsatellites) est avantageux pour notre étude car les allèles peuvent être suivis. Ainsi, nous avons pu évaluer la richesse allélique, les coefficients de parenté et les coefficients de consanguinité.

Les simulations ont ensuite été interprétées par l'approche de Weitzman par le programme WEITZPRO accessible sur site Web (Derban, 2002) et par l'approche de Caballero-Toro (*programme disponible sur demande*). L'analyse des simulations s'est faite à partir d'un échantillonnage de 50 individus pris au hasard dans les populations à taille efficace égale ou supérieure à 50. Pour des populations à taille inférieure la population totale était échantillonnée. Le programme de Caballero-Toro comme celui de Meuwissen n'étant pas accessible librement, nous les avons reprogrammés à partir de leur référence bibliographique (Caballero et Toro, 2002 ; Eding 2002).

Au cours de nos simulations, les populations à l'équilibre ont été mises en place en considérant un nombre de générations au-delà de la valeur  $4Ne$  (quatre fois la taille efficace). Cette valeur a été déterminée comme étant le moment où la population tend vers l'équilibre par l'action de la dérive qui compense celui de la mutation (Malécot, 1948). D'après l'équation (1.12), le taux d'homozygotie moyen ( $\bar{F}$ ) atteint une valeur d'équilibre égale à (Kimura et Crow, 1964)

$$\bar{F} = \frac{1}{1 + 4Ne\beta}, \quad (1.12)$$

où  $Ne$  est la taille efficace et  $\beta$  est le taux de mutation.

Des études antérieures (Laval, 2001) ont évalué un taux d'homozygotie moyen de la population fondatrice chez le porc à 30%. Afin d'obtenir une population à l'équilibre dérive et mutation avec un taux d'homozygotie à 30% sous un modèle IAM, il ne faut pas que la taille efficace soit trop faible ou le taux de mutation trop élevé. Ainsi une taille efficace autour de 300 individus a été considérée pour ce placer à l'équilibre sous IAM.

Nos travaux seront le résultat de simulation sous un modèle de dérive et mutation où la force de dérive a pu être modifiée. La taille efficace des populations, la structure des populations, le taux de mutation, le temps de génération de la population fondatrice et les populations émergentes pouvaient être manipulés. Ce qui nous a permis d'établir des populations à l'équilibre entre dérive et mutation, et des populations sous dérive génétique.

### **2.3 Analyse des populations porcines**

Dans le cadre du projet européen PigBioDiv, 58 populations porcines (*Sus scrofa*) Européennes et 2 populations chinoises composées d'environ 50 individus chacune, ont été génotypées à l'aide de cinquante microsatellites (courtes séquences répétées arrangées en tandem) génétiquement indépendant (au nombre de 2 ou 3 par chromosome). Les microsatellites présentent des avantages non négligeables pour la détermination des liens de parenté et de la structure de populations : les locus peuvent être analysés séparément, ils présentent un polymorphisme élevé et enfin les hétérozygotes peuvent être détectés (marqueurs codominants). Ces données ont été analysées par trois méthodes en génétique conservative. Une méthode développée par Caballero et Toro (2002), la sous-méthode WLM (Weighted Log-linear Model) d'Eding et Meuwissen (2001), et la méthode de Weitzman (1992-1993 ; résultats obtenus par Ollivier et Foulley (2005)).

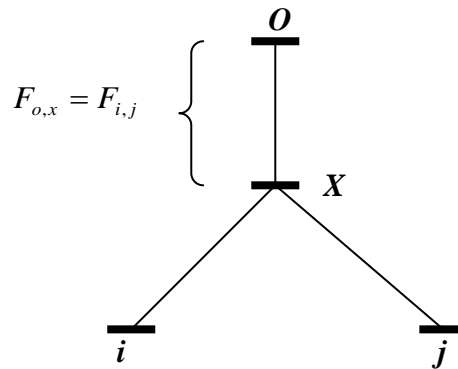
Pour finir, les matrices  $\tilde{f}$  des soixante populations obtenues par chacune des méthodes ont été comparées en construisant des cartes couleurs. Ces cartes représentent par une intensité forte (couleur rouge) un coefficient de parenté ( $f_{i,j}$ ) ou de consanguinité ( $f_{i,i}$ ) élevé et par une intensité faible (couleur jaune) un coefficient faible. La non similarité entre la matrice  $\tilde{f}$  de Caballero-Toro et celle des sous-méthodes d'Eding-Meuwissen, nous ont conduit à comparer les contributions obtenues par la méthode de Caballero-Toro et la sous-méthode WLM.

### III Résultats

#### 3.1 Résolutions mathématiques

##### 3.1.1 Similarité entre la méthode de Caballero-Toro et celle d'Eding-Meuwissen

Imaginons un schéma à deux populations, notées  $i$  et  $j$ , issues d'une population fondatrice notée  $O$  qui a dérivé en  $X$ . Considérons  $p_{i,k}$ ,  $p_{j,k}$ ,  $p_{x,k}$  et  $p_{o,k}$  les fréquences de l'allèle  $k$  dans la population  $i, j, X$  et  $O$  respectivement.



où  $F_{o,x} = F_{i,j}$  est le coefficient moyen de parenté dans la population  $X$ .

L'espérance de  $f_{i,j}$  est alors égale à l'espérance de la somme du produit des fréquences alléliques de la population  $i$  et  $j$  (1.13) :

$$E(f_{i,j}) = E\left(\sum_k p_{i,k} p_{j,k}\right). \quad (1.13)$$

Le produit des fréquences de l'allèle  $k$  dans les populations  $i$  et  $j$  pouvant s'écrire selon l'équation (1.14) en équation (1.15), on obtient l'équation (1.16) en sommant les fréquences :

$$p_{i,k} p_{j,k} = [p_{o,k} + (p_{x,k} - p_{o,k}) + (p_{i,k} - p_{x,k})] * [p_{o,k} + (p_{x,k} - p_{o,k}) + (p_{j,k} - p_{x,k})] \quad (1.14)$$

$$\begin{aligned}
P_{i,k}P_{j,k} &= P_{o,k}^2 + P_{o,k}(P_{x,k} - P_{o,k}) + P_{o,k}(P_{j,k} - P_{x,k}) + P_{o,k}(P_{x,k} - P_{o,k}) + (P_{x,k} - P_{o,k})^2 + \\
&(P_{x,k} - P_{o,k})(P_{j,k} - P_{x,k}) + (P_{i,k} - P_{x,k})P_{o,k} + (P_{i,k} - P_{x,k})(P_{x,k} - P_{o,k}) + (P_{i,k} - P_{x,k})(P_{j,k} - P_{x,k})
\end{aligned} \tag{1.15}$$

$$E\left(\sum_k P_{i,k}P_{j,k}\right) = \sum_k P_{o,k}^2 + E(P_{x,k} - P_{o,k})^2. \tag{1.16}$$

Sachant que la variance  $(E(P_{x,k} - P_{o,k})^2)$  de l'allèle  $k$  entre  $X$  et  $O$  vaut le produit de  $F_{o,x}P_{o,x}(1 - P_{o,k})$  (Laval, 2001) :

$$E(P_{x,k} - P_{o,k})^2 = F_{o,x}P_{o,x}(1 - P_{o,k}), \tag{1.17}$$

nous en déduisons

$$E(f_{i,j}) = E\left(\sum_k P_{i,k}P_{j,k}\right) = \sum_k P_{o,k}^2 + F_{o,x}P_{o,x}(1 - P_{o,k}), \tag{1.18}$$

$$E(1 - f_{i,j}) = (1 - F_{o,x})(1 - \sum_k P_{o,k}^2). \tag{1.19}$$

En pondérant les équations (1.18 et 1.19) par les contributions et en remarquant que  $F_{o,x} = F_{i,j}$ , nous obtenons l'équation (1.20)

$$E\left(\sum_i \sum_j c_i c_j (1 - f_{i,j})\right) = \left(\sum_i \sum_j c_i c_j (1 - F_{i,j})\right) \left(1 - \sum_k P_{o,k}^2\right). \tag{1.20}$$

Comme  $\sum_i c_i = 1$ , cela s'écrit aussi :

$$E\left(\sum_i \sum_j c_i c_j f_{i,j}\right) = \left(\sum_i \sum_j c_i c_j F_{i,j}\right) \left(1 - \sum_k P_{o,k}^2\right) + \sum_k P_{o,k}^2. \tag{1.21}$$

Maximiser la diversité génétique c'est minimiser la fonction (1.21).

- Pour Caballero-Toro, maximiser la fonction de diversité ( $GD$ ) revient à minimiser le premier terme de l'équation (1.21) puisque :

$$GD = 1 - \left(\sum_i \sum_j c_i c_j f_{i,j}\right) = 1 - c' \tilde{f}c. \tag{1.22}$$

Les auteurs négligent le second terme de l'équation (1.21) et ne tiennent pas compte du taux d'homozygotie de la population fondatrice. De plus, ils posent  $\sum_i c_i = 1$  et  $c_i \geq 0$  pour assurer la positivité de toutes les contributions.

- Les sous-méthodes WLM et WLMM de Eding-Meuwissen cherchent à maximiser la diversité génétique :

$$GD = \left( \sum_i \sum_j c_i c_j (1 - F_{i,j}) \right) . \quad (1.23)$$

Ce qui d'après l'équation (1.20) est la même chose que maximiser la diversité selon Caballero-Toro, puisque les deux ne diffèrent que par un facteur multiplicatif (hétérozygotie initiale inconnue). Cependant n'imposant pas les contraintes  $c_i \geq 0$ , les contributions négatives sont mises à zéro.

L'équivalence théorique en espérance entre les deux méthodes et les difficultés numériques rencontrées dans la méthode de Eding-Meuwissen, nous amènent à ne pas prendre en compte cette dernière lors des simulations.

### 3.1.2 *Lien indirect entre les trois méthodes*

L'approche de Caballero-Toro et l'approche d'Eding-Meuwissen sont reliées à la distance de Nei au travers du coefficient de Malécot (1948) et de Wright (1943) sous un modèle IAM.

L'exemple de deux populations avec un pool d'allèles bien définis (voir *Annexe C*) démontre l'égalité entre le coefficient de Malécot ( $J_{XY}$ ) (1948) et la matrice  $f_{i,j}$  de Caballero-Toro. Dans l'équation (1.24) de la distance de Nei standard (1972), Nei précise que  $J_{XY}$  est le coefficient de parenté (Kinship) de Malécot et  $J_X$  ou  $J_Y$  les coefficients de consanguinité (inbreeding) de Wright.

$$D = -\log\left(\frac{J_{XY}}{J_X J_Y}\right) \quad (1.24)$$

$J_{XY}$  est la probabilité de tirer deux allèles identiques, l'un de la population  $X$  et l'autre de la population  $Y$ .  $J_X$  ou  $J_Y$  est la probabilité de tirer deux allèles identiques dans la même population. Il faut noter que cette équivalence n'est vraie que dans le cadre du modèle IAM ; alors le coefficient de parenté représente  $f_{i,j}$  et le coefficient de consanguinité représente  $f_{i,i}$ . Nous pouvons en conclure que l'approche de Caballero-Toro ainsi que celle d'Eding-Meuwissen sont reliées indirectement à la distance de Nei (1972) par le coefficient de Malécot et le coefficient de Wright.

La distance de Nei ayant été utilisée comme distance génétique référentielle lors des calculs des contributions par l'approche de Weitzman (1992-1993), cela nous amène à faire un lien entre les trois approches considérées. Cependant, une des différences majeure reste dans l'évaluation des contributions. La somme des contributions calculées en pourcentage par l'analyse de Weitzman n'est pas de cent pour cent car les contributions sont des pertes de diversité relatives (%), si cette population n'était pas conservée.

### 3.1.3 Schéma en étoile

Lorsque les populations sont toutes issues d'une même population fondatrice  $o$  avec l'absence de population intermédiaire, le modèle est dit en étoile. Dans un schéma en étoile du fait de l'absence de branche en commun entre les populations descendantes, l'équation (1.23) nous donne :

$$GD = 1 - F_1 - \sum_i c_i^2 (F_{i,i} - F_1) , \quad (1.25)$$

où  $F_1 = F_o e^{-2\beta t}$  est la valeur commune de tous les  $F_{i,j} (i \neq j)$ .

Annuler les dérivées de l'équation (1.25) sert à maximiser la fonction de diversité ( $GD$ ). Une contrainte comme estimateur de Lagrange ( $\lambda(\sum_i c_i - 1)$ ) est ajoutée pour assurer que la somme des contributions vaut 1. La dérivée nulle de l'équation (1.25) est représentée par l'équation :

$$\frac{dGD}{dc_i} = 2c_i (F_{i,i} - F_1) + \lambda = 0 . \quad (1.26)$$

Ce qui donne comme solution pour  $c_i$  :

$$c_i = \frac{-\lambda}{2(F_{i,i} - F_1)} \propto \frac{1}{F_{i,i} - F_1} . \quad (1.27)$$

Dans le cas de populations provenant d'une même population fondatrice, nous pouvons en conclure que les contributions sont inversement proportionnelles aux taux d'homozygoties ( $F_{i,i}$ ) diminués de la parenté résiduelle entre populations.

Sous un modèle en IAM  $F_{i,i}$  peut s'écrire selon l'équation (1.28) (Crow et Kimura, 1970).

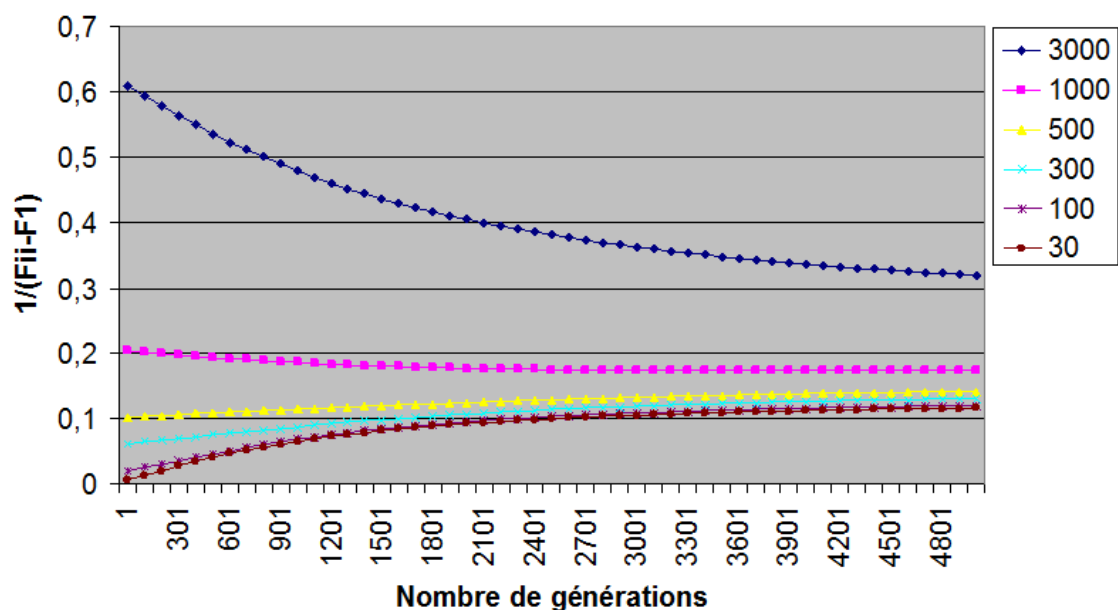
$$F_{i,i} = \frac{1}{1 + 4N_o\beta} e^{-t \frac{1+4Ne\beta}{2Ne}} + \frac{1}{1 + 4Ne\beta} (1 - e^{-t \frac{1+4Ne\beta}{2Ne}}) \quad (1.28)$$

En fixant la taille efficace de la population d'origine ( $N_o$ ) et le taux de mutation ( $\beta$ ), puis en faisant varier la taille efficace des populations observées ( $Ne$ ), une courbe des contributions en fonction du nombre de générations ( $t$ ) a pu être tracée (**Figure 7a**).

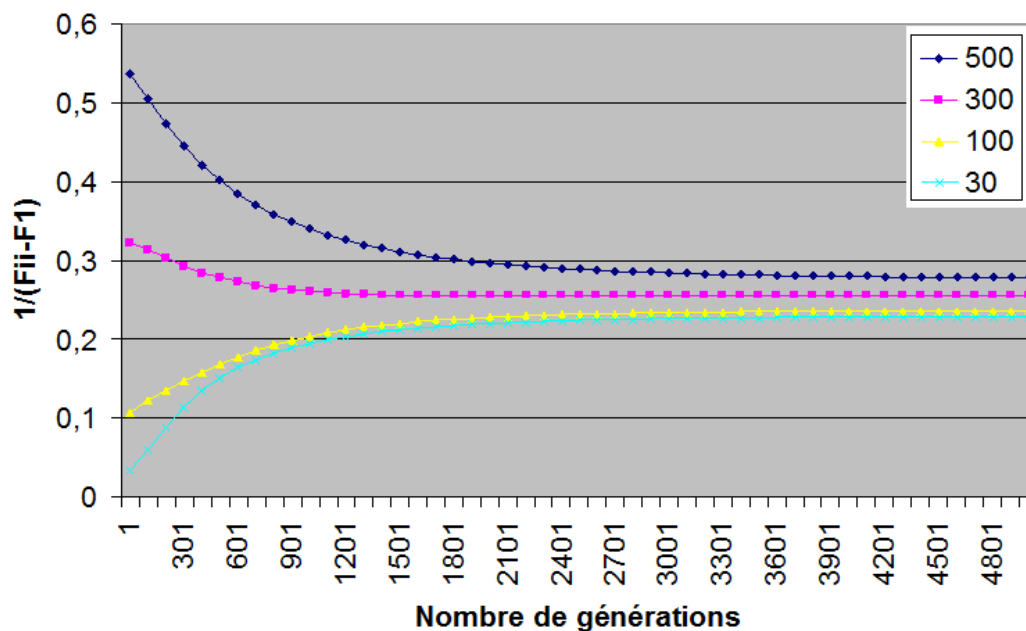
Le taux de mutation a été fixé à  $10^{-4}$  (taux de mutation moyen des microsatellites du porc (Ellegren, 1995)). La taille efficace de la population fondatrice vaut 5834 individus pour être à l'équilibre dérive et mutation avec un taux d'homozygotie de 30%. D'après cette figure, les valeurs des contributions dépendent du moment de l'échantillonnage quelle que soit la taille efficace de la population. C'est à partir d'un certain temps que les contributions se stabilisent car auparavant le nombre de générations depuis la population fondatrice n'est pas assez élevé pour que le nombre d'allèles nouveaux compense l'effet de la dérive génétique.

Cette figure démontre qu'en situation de dérive génétique, la taille de la population est déterminante. Au départ, les contributions dépendent de la taille efficace, puis lorsqu'elles se sont stabilisées, elles dépendent de  $1+4Ne\beta$ .

En comparant les deux figures (**Figure 7a et 7b**) nous pouvons observer que les contributions dépendent également des tailles efficaces des autres populations considérées puisque la somme des contributions est égale à 1.



**Figure 7a.** Evolution des contributions de populations ayant des tailles efficaces allant de 30 à 3000, sachant  $C_i \cong 1/(F_{i,i} - F_1)$  en fonction du nombre de générations. Schéma évolutif : les populations proviennent d'une même population fondatrice à l'équilibre (schéma en étoile).



**Figure 7b.** Evolution des contributions de populations avec des tailles efficaces différentes (de 30 à 500), sachant  $C_i \cong 1/(F_{i,i} - F_1)$  en fonction du nombre de générations, et provenant d'une même population fondatrice à l'équilibre (schéma en étoile).

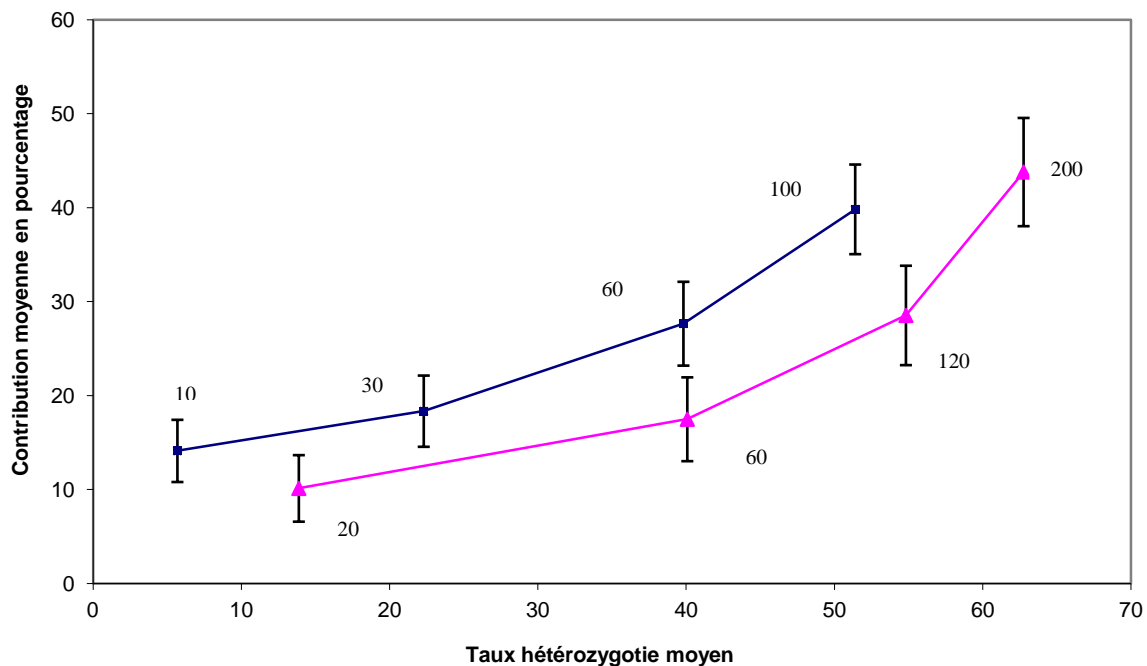


## 3.2 Simulations

### 3.2.1 Comparaison de la méthode de Caballero-Toro et celle de Weitzman

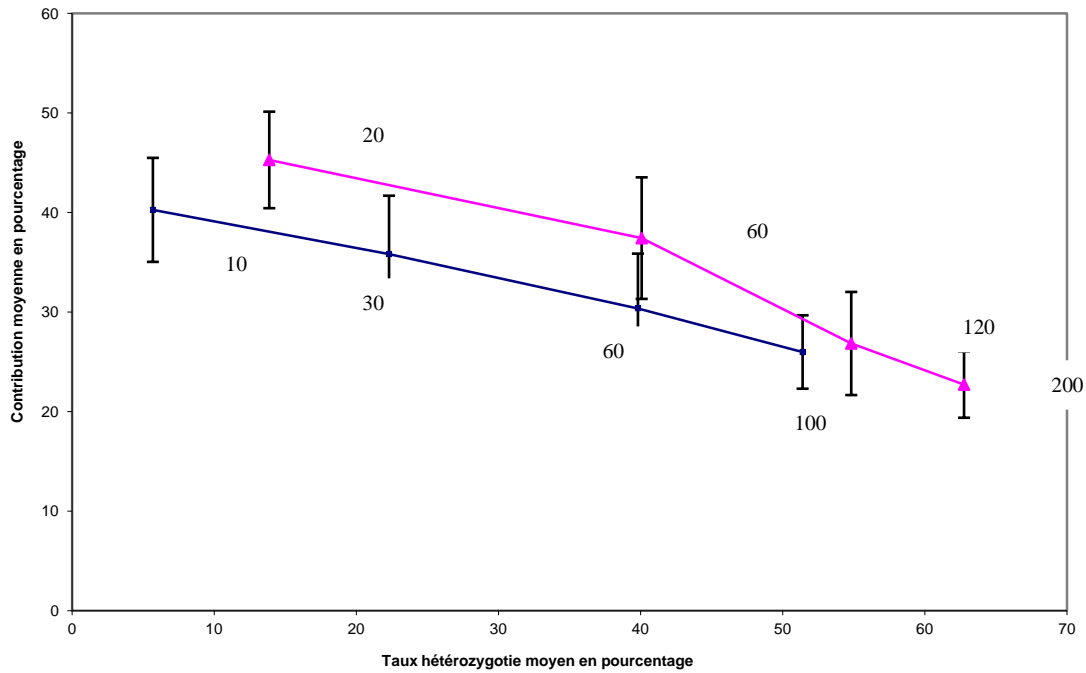
#### 3.2.1.1 Configuration avec des populations sous dérive génétique

Les contributions moyennes, de chacune des 1000 simulations exécutées sur la configuration avec quatre populations sous dérive et la configuration avec trois sous dérive et une à l'équilibre (**Figure 1**), obtenues par l'approche de Caballero-Toro (**Figure 8**) sont corrélées positivement avec le taux d'hétérozygotie moyen. La présence d'une population à l'équilibre ( $N_e=200$ ) parmi l'ensemble des populations entraîne une valeur de sa contribution nettement plus élevée que les autres sous dérive génétique. La valeur des contributions dépend de la distribution des taux d'homozygotie ( $f_{i,i}$ ) des populations.



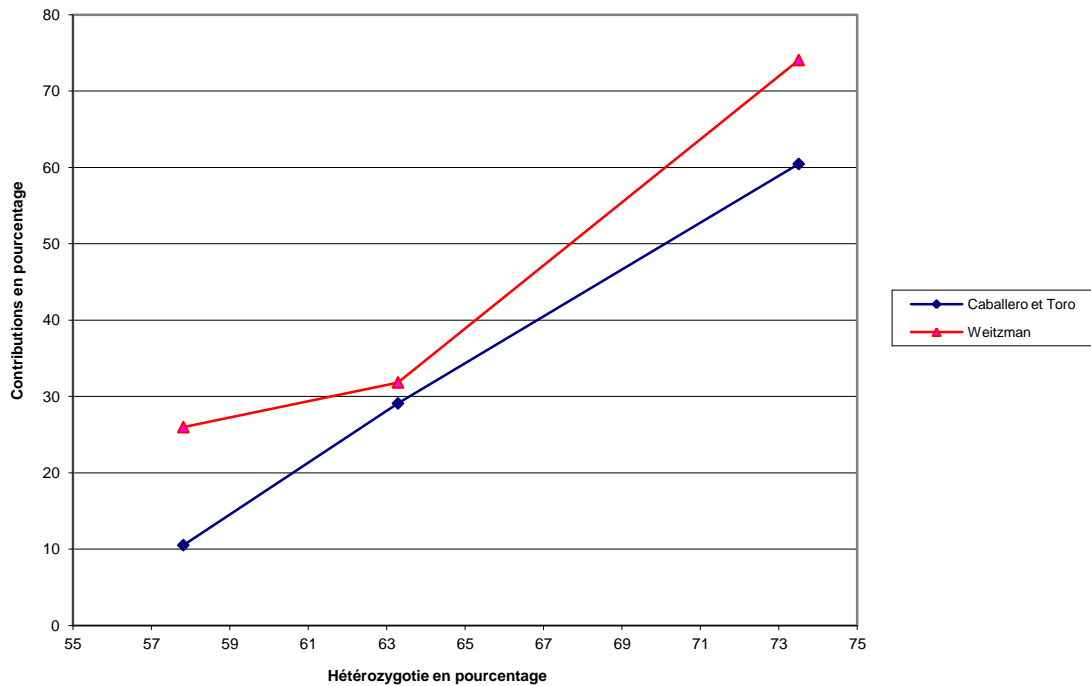
**Figure 8.** Contributions moyennes obtenues sous la méthode de Caballero et Toro (2002) en fonction du taux d'hétérozygotie moyen des mille simulations exécutées pour chaque modèle. L'écart type des contributions moyennes est donné par un trait plein. Le doublement de la taille efficace ( $N_e$ ) des populations du premier modèle (courbe bleue) au second (courbe rose) est donné par les valeurs des ' $N_e$ ' à l'emplacement de chaque point.

Cette observation s'oppose à l'approche de Weitzman (**Figure 9**) dont les contributions moyennes ont des valeurs plus élevées lorsque les populations sont sous dérive génétique ( $N_e=10, 30, 60$  et  $100$  dans le premier modèle et  $N_e=20, 60$  et  $120$  dans le second modèle). Nous pouvons en conclure que les contributions sont corrélées négativement avec le taux d'hétérozygotie. Les populations avec une valeur de  $f_{i,i}$  élevée auront une contribution forte.



**Figure 9.** Contributions moyennes obtenues sous la méthode de Weitzman (1992-93) en fonction du taux d'hétérozygotie moyen des mille simulations exécutées pour chaque point. L'écart type des contributions moyennes est donné par un trait plein. Le doublement de la taille efficace ( $N_e$ ) des populations du premier modèle (courbe bleue) au second (courbe rose) est donné par les valeurs des ' $N_e$ ' à l'emplacement de chaque point.

### 3.2.1.2 Configuration avec l'ensemble des populations à l'équilibre



**Figure 10.** Contributions en pourcentage obtenues par l'approche de Caballero et Toro (courbe bleue) et de Weitzman (courbe rouge) en fonction des taux d'hétérozygoties de populations à l'équilibre dérive et mutation (Voir Annexe C.4). Les simulations ont été obtenues à partir de 50 locus de 400 allèles fondateurs.

Dans le cas d'un modèle avec des populations à l'équilibre dérive et mutation (**Figure 2**), l'approche de Weitzman et de Caballero-Toro sont identiques. Elles attribuent une contribution maximale à la population ayant un taux d'hétérozygotie le plus élevé (**Figure 10**).

### 3.2.2 Effet du nombre de populations

#### 3.2.2.1 Populations sous dérive génétique (**Figure 3**)

Quelle que soit la simulation sur les 1000 analysées par l'approche de Caballero-Toro, un nombre de populations excédant 14 laisse apparaître des contributions nulles (**Tableau 2** ; voir **Annexe D**). Les coefficients de corrélation entre le taux d'hétérozygotie et les contributions obtenues, en faisant varier le nombre de populations de 8 à 24, sont positives mais faibles lorsque le nombre de populations augmente (**Tableau 2**). Un nombre de populations trop important (supérieur à 14 populations) entraîne des contributions qui ne dépendent plus du taux d'homozygotie ( $f_{i,i}$ ). Cette affirmation s'applique dans le cas où toutes les populations, ont une taille efficace faible et semblable, subissent une même force de dérive génétique, et sont proches génétiquement. Ces résultats s'expliqueraient par les valeurs élevées (3,1% à 3,87%) des écarts types moyens sur 1000 simulations (voir **Annexe E**).

L'analyse de la même simulation par l'approche de Weitzman ne donne jamais des contributions nulles (**Tableau 2** et voir **Annexe F**). Néanmoins les coefficients de corrélation sont faibles (**Tableau 2**). Dans le cas où toutes les populations ont la même taille efficace, subissent la même force de dérive génétique, et sont proches génétiquement, les contributions ne dépendent plus du taux d'homozygotie.

#### 3.2.2.2 Populations à l'équilibre dérive et mutation (**Figure 4**)

Lorsque les populations sont à l'équilibre, l'approche de Caballero-Toro (voir **Annexe G**) montrent des contributions d'autant plus corrélées positivement aux taux d'hétérozygotie que le nombre de populations est extrême (24 populations ou 8 populations) (**Tableau 2**). Les écarts types moyens observés sur 1000 simulations diminuent de 1,6% pour 8 populations à 0,43-0,46% pour 24 populations (voir **Annexe H**).

Dans l'approche de Weitzman la distance génétique  $d(i, j)$  entre les populations à l'équilibre est la distance de Reynolds ( $D_R$ ) (1983) et s'écrit selon l'équation (1.29).

Population sous dérive génétique						Population à l'équilibre						Nombre de Populations	
Caballero-Toro			Weitzman			Caballero-Toro			Weitzman				Coefficient de corrélation
8	14	24	8	14	24	8	14	24	8	14	24		
0,75	0,13	0,25	-0,26	-0,57	-0,14	0,89	0,69	0,84	-0,90	-0,95	-0,96		
0	1	5	0	0	0	0	0	0	0	0	0		

**Tableau 2.** Nombre de contributions nulles et coefficients de corrélation entre le taux d'hétérozygotie et les contributions obtenues par l'analyse de Caballero et Toro et celle de Weitzman. Ces résultats proviennent de l'analyse des configurations plaçant les populations sous dérive génétique et à l'équilibre dérive et mutation. Pour chaque type de configuration le nombre de populations variés de 8 à 24 populations.

$$D_R = \frac{1}{2} \frac{\sum_k (p_{i,k} - p_{j,k})^2}{1 - \sum_k (p_{i,k} p_{j,k})} \quad (1.29)$$

Comme la distance de Nei standard, la distance de Reynolds prend en compte les fréquences de l'allèle  $k$  de la population  $i$  et  $j$ . La distance de Nei n'a pu être utilisée à cause de l'absence d'allèles communs entre les populations. En effet, pour obtenir des populations à l'équilibre, le nombre de générations, entre la population fondatrice et les populations descendantes, s'élève à  $4Ne$ . Sous IAM, les populations se sont retrouvées avec un coefficient de parenté nul. Le logarithme d'une valeur nulle n'étant pas valable, la distance de Nei standard devient incorrecte et ne peut pas être utilisée comme distance génétique. Néanmoins l'analyse des populations à l'équilibre montre une corrélation négative entre le taux d'hétérozygotie et les contributions. Cette corrélation est d'autant plus négative que le nombre de populations à l'équilibre augmente (*Tableau 2* et voir *Annexe I*). Ces résultats s'expliqueraient par l'augmentation des écarts types moyens des 1000 simulations allant de 3 populations (1,76-1,78%) à 24 populations (2,09-2,21%) (Voir *Annexe J*). Bien que l'analyse par l'approche de Weitzman produise des corrélations négatives entre le taux d'hétérozygotie et les contributions, les valeurs absolues des corrélations sont supérieures à celle de Caballero-Toro.

### 3.2.3 Effet du coefficient de parenté

Dans le cas du premier modèle (*Figure 5*), l'approche de Caballero-Toro et de Weitzman va attribuer des contributions faibles lorsque les populations, (2) et (3), sont proches génétiquement (*Tableau 3*). Les contributions pour les deux approches ne dépendent plus du taux d'hétérozygotie. Dans le second modèle (*Figure 6*), les deux approches attribuent des contributions en fonction du taux d'hétérozygotie des populations (*Tableau 4*).

Nous pouvons en conclure deux points cruciaux. Un coefficient de parenté élevé entre deux populations entraîne des contributions faibles en dépit du taux d'hétérozygotie ou d'homozygotie des populations. Un coefficient de parenté faible entre deux populations entraîne des contributions qui dépendront du taux d'hétérozygotie ou homozygotie de ces populations. Ces résultats viennent rejoindre les résolutions mathématiques du modèle en étoile. Lorsque les populations proviennent d'une même population fondatrice et si elles ont divergé assez longtemps pour être génétiquement différentes, les contributions seront proportionnelles à l'inverse du taux d'homozygotie.

Populations	Richesse allélique total	Richesse allélique échantillon	Taux Hétérozygotie (%)	Contributions (%) par Caballero-Toro	Contributions (%) par Weitzman
1	431	368	65	62	92
2	305	259	50	24	13
3	296	261	53,5	14	8

**Table 3.** Contributions obtenues par l'approche de Caballero et Toro (2002) et l'approche de Weitzman (1992-1993) dans le cas du modèle Figure 5. Simulations avec 50 locus à partir de 400 allèles fondateurs.

Populations	Richesse allélique total	Richesse allélique échantillon	Taux Hétérozygotie (%)	Contributions (%) par Caballero-Toro	Contributions (%) par Weitzman
1	128	117	17	35	59
2	132	114	17,22	32,8	52
3	134	123	16,13	31,2	41

**Table 4.** Contributions obtenues par l'approche de Caballero et Toro (2002) et l'approche de Weitzman (1992-1993) dans le cas du modèle Figure 6. Simulations avec 50 locus à partir de 400 allèles fondateurs.

La richesse allélique ( $R$ ) évaluée au sein des deux derniers modèles de simulations (**Figure 5** et **6**) vient compléter notre compréhension sur la relation des contributions avec la maximisation de la diversité génétique (**Tableau 3** et **4**). En regardant les résultats obtenus dans le **Tableau 4**, nous pouvons observer que la population, avec la richesse allélique la plus élevée ( $R= 134$  dans la population totale et  $R= 123$  dans l'échantillon), est la population avec la contribution la plus faible. Dans ce cas de figure, l'évaluation de la diversité génétique par les deux approches ne semble pas cohérente en terme de maximisation de la variabilité génétique.

### 3.3 Analyse des populations porcines

L'analyse des soixante populations porcines européenne par l'approche de Caballero-Toro (**Tableau 5**) donne une majorité de contributions nulles. Au total, treize populations sur les soixante ont une contribution supérieure à zéro. La population créole (FRCR) détient la contribution la plus élevée. Les deux populations de Meishan (GBMS et FRMS), la Landrace Islandaise (ISLR) et la Casertana (ITCS) possèdent des contributions également supérieures à l'ensemble des populations. Parmi les autres populations ayant une contribution non nulle à valeur peu élevée, on trouve Negro Canario (ESNC), la Landrace Danoise parmi les douze Landrace, la Negra Iberica (ESNI), la Manchada de Jabugo, deux populations Large White (une française (FRLW) et une allemande (DELW)) parmi les dix Large White, la Hampshire Britannique (GBHA) sur les deux races de Hampshire. Cependant ces populations ne correspondent pas aux populations avec les taux d'hétérozygoties les plus élevés. Certaines populations parmi les 47 restantes ont un taux d'hétérozygotie plus élevé que les populations avec des contributions non nulles. C'est le cas de la Nera Siciliana (ITNS). Ces observations viendront appuyer les résultats obtenus lors des simulations sous dérive génétique.

Concernant l'approche de Weitzman (**Tableau 6**), malgré l'absence de contribution nulle, la distribution des contributions semble aléatoire.

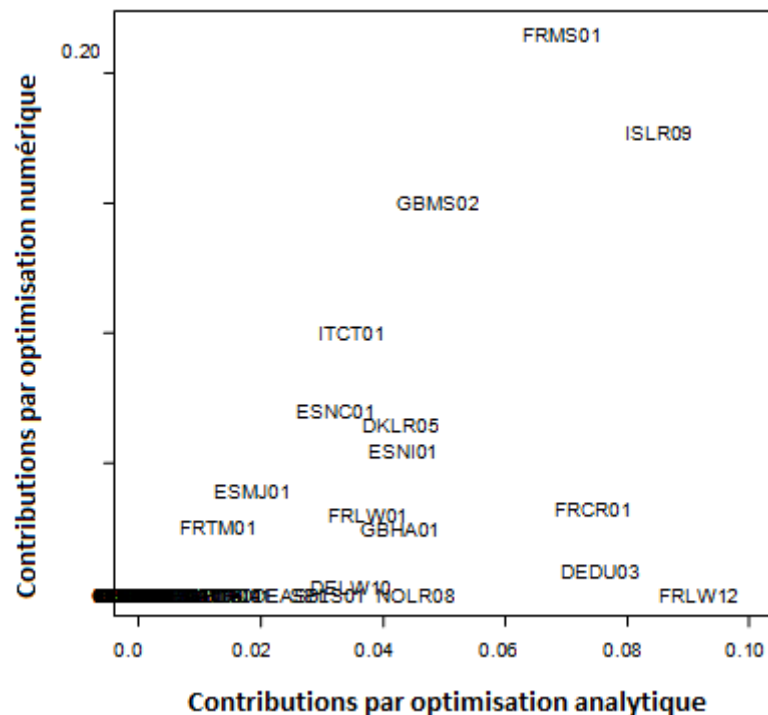
Les résolutions mathématiques ont démontré la similarité en espérance entre la méthode de Caballero-Toro et celle d'Eding-Meuwissen. Cependant les contributions des deux méthodes sur les soixante populations porcines ne sont pas exactement similaires (**Figure 11**). Il semblerait que l'évaluation de la matrice  $\tilde{f}$  par optimisation numérique (Caballero-Toro) ou analytique (Eding-Meuwissen) donne des résultats différents quand le nombre de population est élevé. Les contributions nulles sont moins nombreuses par la méthode d'Eding-Meuwissen.

Populations	Taux hétérozygotie (%)	Contributions (%)	Populations	Taux hétérozygotie (%)	Contributions (%)
1 DEAS01	62,03	0	31 GBLB01	51,46	0
2 GBBK01	43,62	0	32 GBLW05	57,34	0
3 PTBI01	63,8	0	33 GBLW06	55,66	0
4 GBL01	56,99	0	34 GBLW07	48,95	0
5 GBBS01	61,74	0	35 FRLW01	60,13	3,07
6 DEBB01	57,44	0	36 FRLW08	52,06	0
7 ITCA01	48,15	0	37 FRLW09	55,33	0
8 ITCT01	60,14	10,03	38 FRLW12	60,31	0
9 ITCS01	46,43	0	39 ITLW03	59,88	0
10 FRCR01	67,3	32,8	40 DELW02	60,74	0
11 FRDR01	59,83	0	41 DELW10	51,73	2,9
12 GBDU02	48,4	0	42 GBLE01	57,54	0
13 ITDU01	53,88	0	43 SELS01	57,16	0
14 DEDU03	52,09	9	44 ESMJ01	46,15	3,95
15 GBGO01	44,45	0	45 DEMA01	46,76	0
16 GBHA01	48,44	2,5	46 GBMS02	55,4	15,09
17 DEHA02	53,58	0	47 FRMS01	58,4	21,49
18 FRLA01	54,15	0	48 GBMW01	51,99	0
19 GBLR10	54,79	0	49 ESNC01	50,58	7,04
20 GBLR11	54,52	0	50 ESNI01	57,59	5,52
21 GBLR12	54,11	0	51 ITNS01	66,33	0
22 FRLR01	60,23	0	52 GBPI04	54,43	0
23 FRLR13	57,23	0	53 FRPI02	57,28	0
24 ITLR03	61,17	0	54 FRPI05	50,43	0
25 DELR14	56,57	0	55 DEPI03	57,18	0
26 DKLR04	54,13	0	56 CZPR01	64,51	0
27 DKLR05	57,34	6,5	57 PLPU01	56,02	0
28 FILR06	53,86	0	58 ESRE01	57,93	0
29 NOLR08	53,72	0	59 GBTA01	45,91	0
30 ISLR09	67,67	17,74	60 FRTM01	64,01	2,57

**Tableau 5.** Contributions en pourcentage et taux d'hétérozygotie obtenu pour chacune des soixante populations porcines par l'approche de Caballero et Toro (2002).



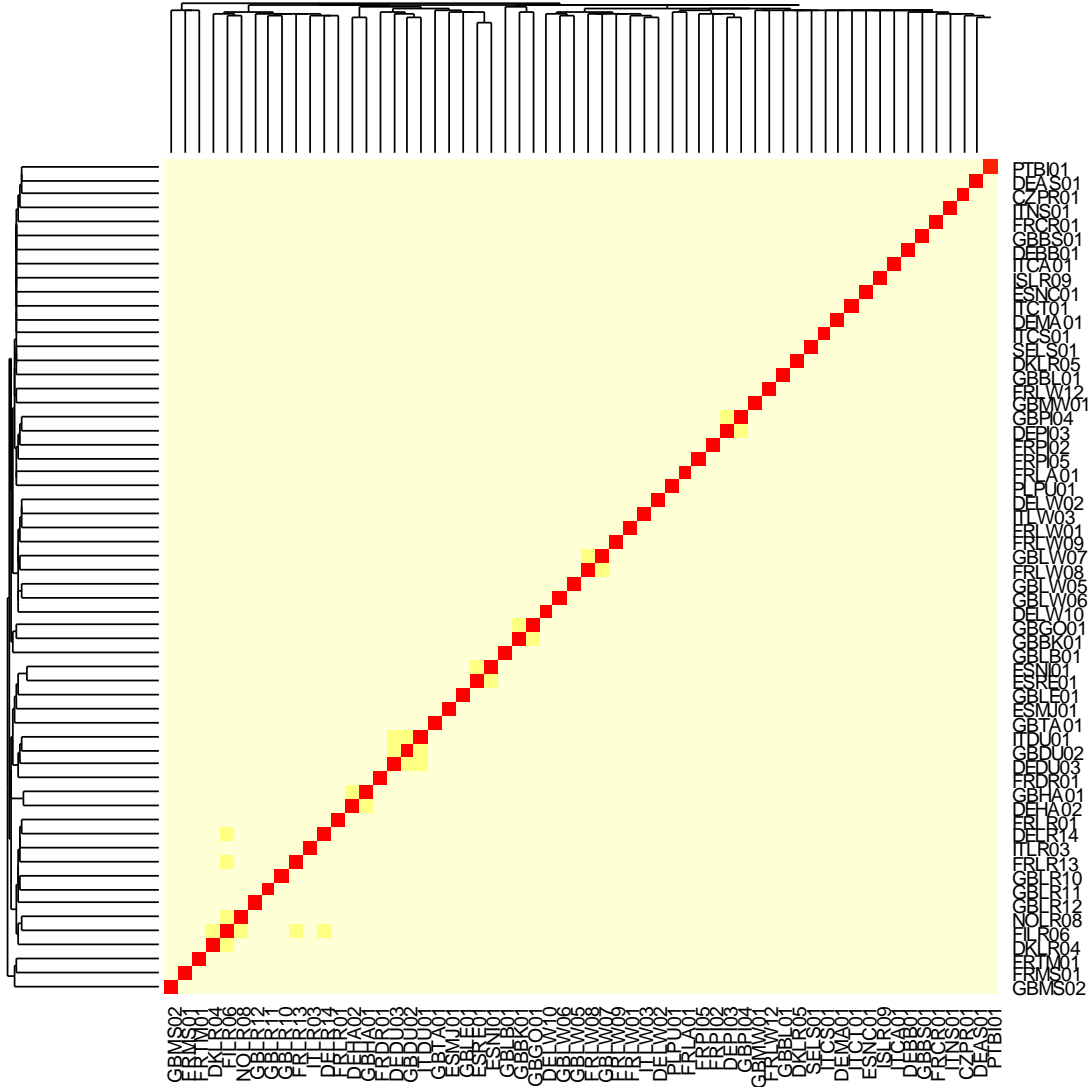
Nous avons donc cherché à comparer les matrices  $\tilde{f}$  des trois méthodes. Tout d'abord, la matrice de la sous-méthode WLMM (*Figure 12*) sous-estime fortement les coefficients de parenté ( $f_{i,j}$ ) entre les populations. Les coefficients de consanguinité ( $f_{i,i}$ ) sont surreprésentés. Les autres matrices montrent une hétérogénéité dans les deux coefficients. La matrice  $\tilde{f}$  WLM (*Figure 13*) est quelque peu différente de la matrice  $\tilde{f}$  ULM (*Figure 14*) ou de la matrice  $\tilde{f}$  de Caballero-Toro (*Figure 15*). Elle possède des coefficients de parenté intergroupe moins importante qu'intragroupe. La matrice  $\tilde{f}$  ULM (*Figure 14*) est proche de la matrice  $\tilde{f}$  de Caballero-Toro (*Figure 15*). Lorsque le nombre de populations devient élevé, les évaluations des coefficients de parenté ( $f_{i,j}$ ) et de consanguinité ( $f_{i,i}$ ) sont différentes et entraînent des contributions différentes (*Figure 11*).



**Figure 11.** Contributions par optimisation numérique (approche de Caballero et Toro) en fonction des contributions par optimisation analytique (approche de Eding et Meuwissen : WLM (Weighted Log-linear Model)) des soixante populations porcines européennes. Sur ce graphique peu de populations ont des contributions non nulles : Meishan (MS), Landrace (LR), Casertana (CT), Tia meslan (TM), Negro canario (NC), Negra iberica (NI), Hampshire (HA), Créole (CR), Manchada de Jabugo (MJ) et Duroc (DU).

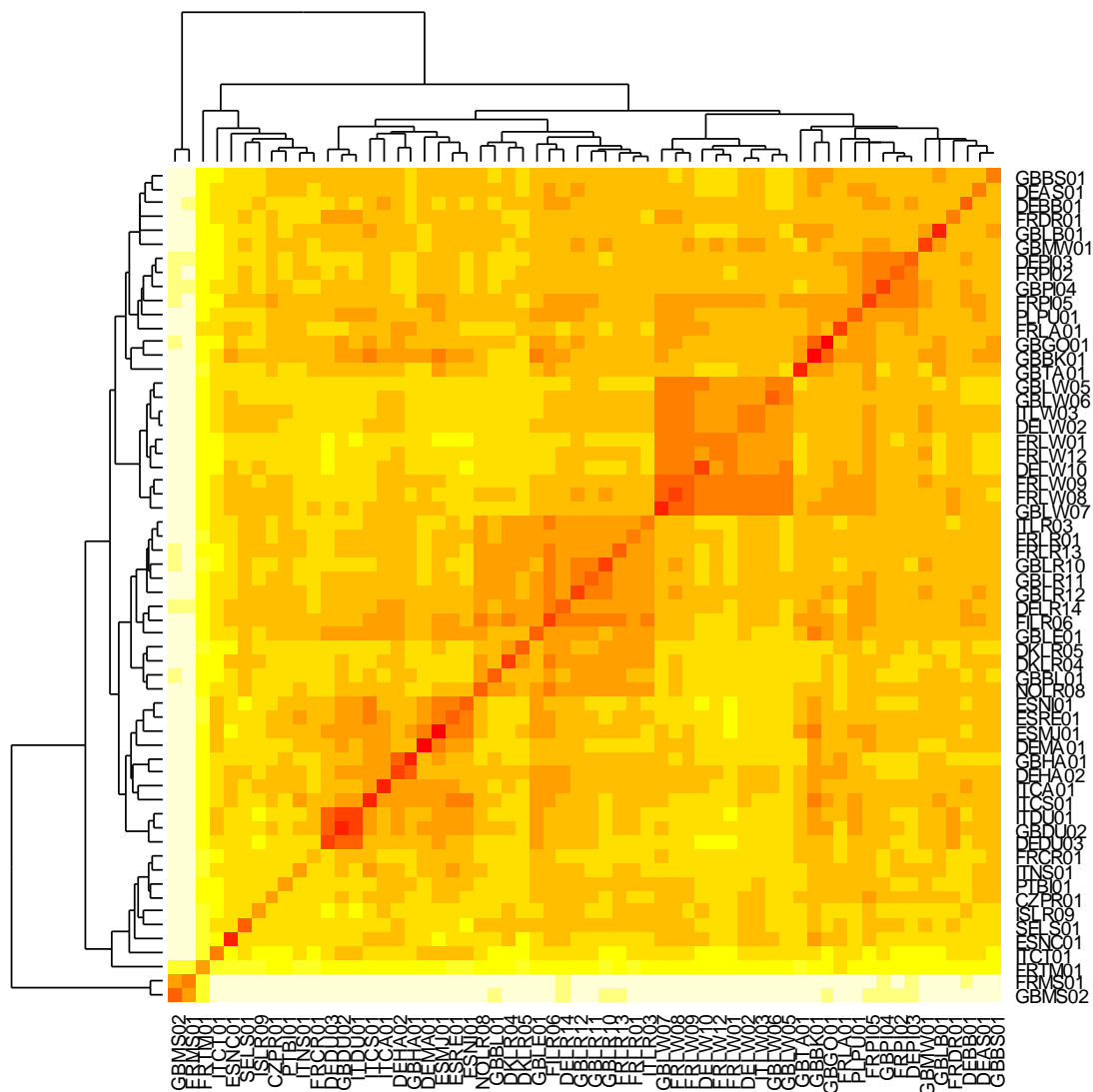
<b>Populations porcines européennes</b>	<b>Contributions en %</b>
CZPR01 (Presticke)	0.48
DEAS01 (Angler Sattelschein)	0.98
DEBB01 (Bunte Benheimer)	2.00
DEDU03 (Duroc)	1.54
DEHA02 (Hampshire)	1.07
DELR14 (German Landrace)	1.32
DELR02 (German Large White)	0.47
DELR10 (German Large White)	1.82
DEMA01 (Mangalica)	3.01
DEPI03 (German Piétrain)	0.78
DKLR04 (Danish Landrace)	0.85
DKLR05 (Danish Landrace)	1.04
ESNC01 (Negro Canario)	3.19
ESNI01 (Negra Iberica)	0.56
ESMJ01 (Manchada de Jabugo)	2.60
ESRE01 (Retinta)	0.04
FILR06 (Finnish Landrace)	0.95
FRCR01 (Créole Française)	0.58
FRDR01 (Scapaag)	1.68
FRLA01 (Laconie)	1.26
FRLR01 (Landrace Française)	0.32
FRLR13 (Landrace Française)	0.91
FRLW01 (Large White Française)	0.73
FRLW08 (Large White Française)	1.21
FRLW09 (Large White Française)	0.89
FRLW12 (Large White Française)	0.28
FRPI02 (Piétrain Française)	0.23
FRPI05 (Piétrain Française)	1.58
FRTM01 (Tai meslane)	2.34
GBBK01 (Berk Shire)	2.86
GBBL01 (British Lop)	1.62
GBBS01 (British Saddleback)	1.05
GBDU02 (Duroc)	0.97
GBGO01 Gloucestershire Old Spots)	2.59
GBHA01 (Hampshire)	1.67
GBLB01 (Large Black)	2.40
GBLE01 (Leicoma)	0.78
GBLR10 (British Landrace)	0.85
GBLR11 (British Landrace)	1.24
GBLR12 (British Landrace)	0.79
GBLW05 (British Large White)	0.90
GBLW06 (British Large White)	0.80
GBLW07 (British Large White)	1.60
GBMW01 (British Middle White)	2.45
GBPI04 (British Piétrain)	1.38
GBTAA01 (Tamworth)	3.27
ISLR09 (Irish Landrace)	1.29
ITCA01 (Calabrese)	2.75
ITCS01 (Cinta Cinese)	2.84
ITCT01 (Casertana)	2.32
ITDU01 (Italian Duroc)	0.25
ITLR03 (Italian Landrace)	0.49
ITLW03 (Italian Large White)	0.35
ITNS01 (Nera Siciliana)	1.12
NOLR08 (Norwegian Landrace)	1.37
PLPU01 (Pulawska Spot)	1.97
PTBI01 (Bisaro)	1.43
SELS01 (Lindrödssvin)	1.89

**Tableau 6.** Contributions par population en pourcentage obtenues par l'approche de Weitzman (1992-1993) (Ollivier et Foulley, 2005).



**Figure 12.** Représentation de la matrice  $\tilde{f}$  obtenue avec la sous sous-méthode WLMM (Weighted Log-linear Mixed Model) d’Eding et Meuwissen (2002) avec la variance résiduelle optimale égale à 95 000.





**Figure 14.** Représentation de la matrice  $\tilde{f}$  obtenue par la sous-méthode ULM (Unweighted Log-linear Model) d’Eding et Meuwissen (2002).

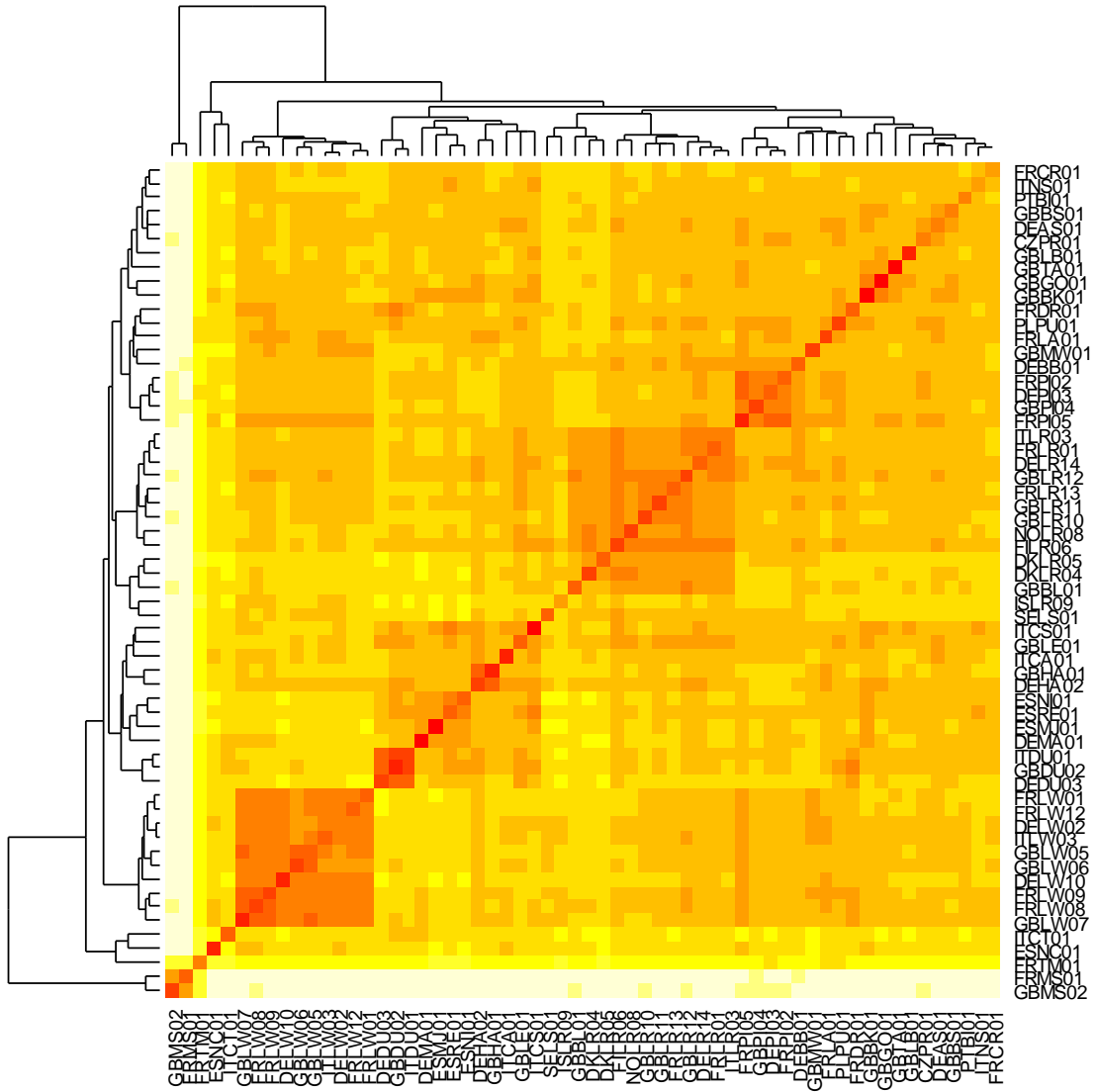


Figure 15. Représentation de la matrice  $\tilde{f}$  obtenue par l'approche de Caballero et Toro (2002).

## IV Discussion

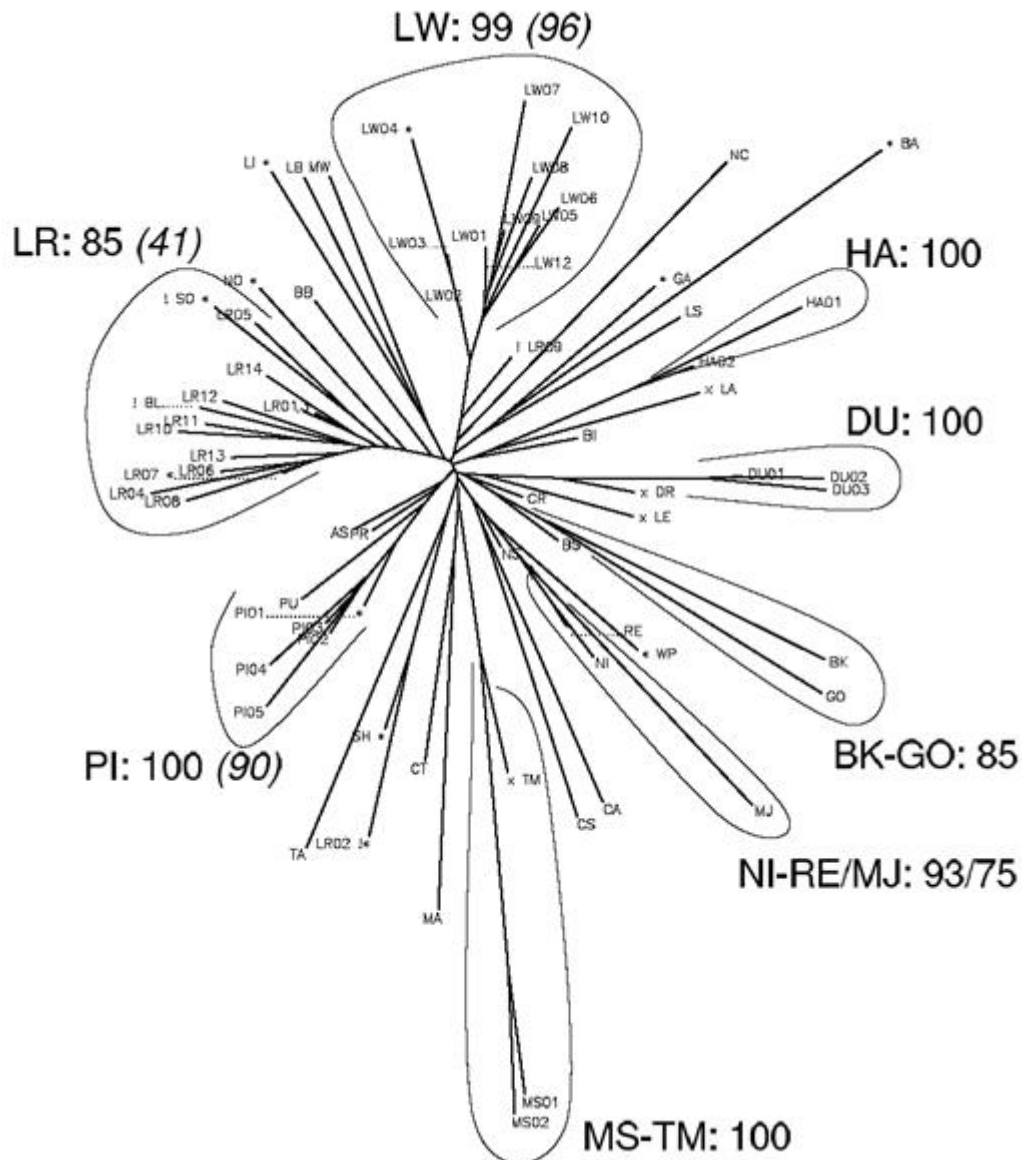
Les décisions de conservation reposent sur la structure et la dynamique des populations. Dans le cas de grandes populations à l'équilibre entre dérive et mutation, les trois approches sont équivalentes et donnent des résultats satisfaisants. C'était dans ce cadre que Weitzman avait proposé sa méthode, avant qu'elle ne soit utilisée par d'autres auteurs dans des contextes différents. Néanmoins cette conclusion n'est pas valable pour un nombre élevé de populations faiblement apparentées et avec la même taille efficace. Les contributions ne sont plus proportionnelles aux taux d'hétérozygoties. Les contributions évaluées par la méthode de Weitzman deviennent proportionnelles aux taux d'homozygotie. Cette méthode conservera des populations consanguines. Parce qu'une population consanguine verra sa variabilité génétique appauvrie, cette approche ne peut être retenue en vue de la conservation de l'espèce. Ce phénomène peut s'expliquer par la présence de dérive génétique plus ou moins faible au sein des populations à l'équilibre. Cette dérive entraînerait la fixation ou l'élimination de certain(s) allèle(s), diminuant le taux d'hétérozygotie et le lien de parenté avec les autres populations. Ainsi la population avec une longueur de branche plus longue, verra sa contribution augmenter.

Dans le cas d'un nombre de populations faible sous dérive génétique, la méthode de Caballero-Toro et celle d'Eding-Meuwissen donnent des résultats opposés à la méthode de Weitzman. La méthode de Weitzman entraîne des attributions de contributions proportionnelles aux taux d'homozygoties. Cette méthode se comporte comme dans le cas des analyses sur un grand nombre de populations à l'équilibre ayant des liens de parentés faibles et des tailles efficaces identiques. Cependant, lorsque le nombre de populations sous dérive génétique devient élevé, des contributions nulles apparaissent avec l'analyse par Caballero-Toro ce qui provoque des corrélations positives entre les contributions qui diminuent les taux d'hétérozygoties. Pour Weitzman, la corrélation négative qui existait entre les taux d'hétérozygoties et les contributions devient très faible.

Les simulations ont permis de démontrer l'importance des liens de parenté et du nombre de populations sur l'attribution des contributions. Des liens de parenté importants entre de nombreuses populations sous dérive génétique entraînent des contributions qui ne sont plus proportionnelles aux taux d'hétérozygoties.

Concernant la méthode d'Eding-Meuwissen, elle est équivalente en espérance à la méthode de Caballero-Toro. Les différences proviennent de l'estimation de la matrice symétrique  $\tilde{f}$  et de

la procédure d'optimisation. Il semblerait que les différences entre les matrices soient dues en partie au grand nombre de populations. Le choix de la méthode d'estimation de la matrice  $\tilde{f}$  reste difficile car nous sommes confrontés au grand nombre de contributions nulles. Une estimation alternative serait de calculer les  $F_{i,j}$  à partir des longueurs de branche d'un arbre NJ (Neighbor-Joining Tree ; Saitou et Nei, 1987) établi avec des distances de Reynolds (1983).



**Figure 16.** Arbre NJ (Neighbor-Joining tree) (Saitou et Nei, 1987) à partir de la distance de Reynolds (1983) des soixante populations porcines européennes (SanCristobal et al. 2005). Sur cet arbre les grandes races ont été regroupées : Duroc (DU), Hampshire (HA), Large White (LW), Meishan (MS), Piétrain (PI) et Landrace (LR). Ainsi que certaines populations locales.



En regardant l'arbre NJ des soixante populations porcines, il n'y a pas de raison apparente d'obtenir des contributions nulles à l'intérieur des groupes (**Figure 16** ; SanCristobal *et al.* 2005). Il est à noter que la matrice de la sous-méthode WLMM (Weighted Log-linear Model) n'est pas satisfaisante car elle ignore la parenté entre les populations. Or, nos analyses sur la parenté montrent les liens forts existants entre certaines grandes races commerciales (Large White, Duroc, Hampshire (HA), Landrace (LR), Meishan (MS)) et celles de lignées synthétiques (Tia meslan (FRTM) et Scapaag (FRDR)) (Nikolic *et al.* 2009). Comme la méthode d'Eding-Meuwissen, la méthode de Caballero-Toro va choisir à l'intérieur des races une ou deux populations seulement sans considérer l'ensemble du groupe fortement lié.

L'analyse des populations porcines par l'approche de Weitzman donne des contributions qui ne sont ni proportionnelles aux taux d'hétérozygotie, ni proportionnelles aux taux d'homozygotie. Ceci s'explique par son fonctionnement fondé sur les différences entre les populations. Quel que soit le taux d'homozygotie des populations, cette méthode attribue des contributions en fonction de la parenté entre les populations ( $f_{i,j}$ ). Les incohérences apparaissent lorsque les populations sont proches génétiquement. La méthode de Weitzman pourrait être utilisée avec précaution sur des populations à l'équilibre dérive et mutation ayant une parenté faible.

Un autre paramètre décisif dans l'évaluation des contributions est le temps d'échantillonnage. Au sein d'un schéma en étoile, les contributions diffèrent d'un temps à un autre. Les contributions dépendent de la taille efficace de la population et du temps d'échantillonnage. On peut penser à une situation similaire dans le cas des populations porcines. Sachant que les populations porcines ont divergé depuis peu d'une même population fondatrice, les contributions ne sont donc pas stabilisées et varient d'un temps à un autre. En plus de la structure et la dynamique des populations, il semble important de considérer le moment de l'échantillonnage pour choisir au mieux la méthode d'analyse génétique.

La richesse allélique n'est pas un paramètre considéré par les trois méthodes. Cependant elle peut être évaluée pour avoir une idée de la diversité génétique. Lors des dernières simulations (**Figure 5 et 6**), nous avons observé qu'une population avec la richesse allélique la plus élevée se retrouvait avec la contribution la plus faible par la méthode de Caballero-Toro et de Weitzman (**Tableau 4**). Nous pouvons donc nous poser la question sur les analyses de la diversité génétique s'appuyant seulement sur les fréquences alléliques.

Ces travaux font émerger la problématique du choix de la méthode dans le cas de populations subissant de la dérive génétique. Diverses critiques ont vu le jour ces dernières

années sur les mesures d'évaluation de la diversité génétique. Mainwaring (2001) critique les projets de conservation basés sur les informations génétiques. Il considère que la mesure de diversité fondée sur les différences génétiques nécessite plus d'information que nous avons à notre disposition. Même si l'information génétique disponible est suffisante, la question est de savoir si la diversité ainsi mesurée fournit des valeurs significatives de la diversité. Goudy (1997) vient appuyer ces critiques en affirmant que nous ne pouvons rattacher une quelconque valeur à une ressource essentielle. Le thème de la biodiversité est un sujet très complexe car il mêle divers concepts et théories. La biodiversité en termes de diversité génétique reste à ce jour mal définie. Bien que nos travaux aient cherché une méthode efficace pour maximiser la diversité génétique au sein de l'espèce *Sus scrofa*, nous ne pouvons pas ignorer l'inexistence de valeurs formelles pouvant la quantifier. Il apparaît que nous ne puissions amener des affirmations par les fréquences génétiques uniquement. Les informations contenues dans le matériel génétique sont nombreuses et compliquent l'interprétation.

## V Conclusion

La comparaison de trois méthodes d'évaluation et de maximisation de la diversité génétique laisse entrevoir leurs limites en fonction de la structure et de la dynamique des populations. Ces méthodes couramment utilisées dans les projets de conservation d'animaux d'élevage sont des outils très intéressants mais il semble qu'ils ne soient pas adaptés en présence d'un grand nombre de populations sous dérive génétique ayant des individus reproducteurs fortement apparentés. Les degrés de parenté et de consanguinité ont d'importantes conséquences sur l'évaluation et la maximisation de la diversité génétique.

## VI Références

- Binder, E., 1967. La génétique des populations. Presse universitaire de France
- Bureau des Ressources génétiques. 2005.  
<http://www.brg.prd.fr/>
- Caballero, A., Toro M. A., 2002. Analysis of genetic diversity for the management of conserved subdivided populations, *Conservation Genetics*, 3, 289-299
- Chevalet, C., Rochambeau, H., 1985. Predicting the genetic drift in small populations. *Livestock Production Science*, 13: 207-218
- Derban, S., Foulley J-L., Ollivier L., 2002. WEITZPRO: a software for analysing genetic diversity. INRA, Paris (<http://www.sgqa.jouy.inra.fr/diffusions.htm>)
- Eding, H., Meuwissen T. H. E., 2002. Estimation of marker based kinships to construct core sets for gene banks, PhD, *Conservation of Genetic Diversity: Assessing genetic variation using marker estimated kinship*, 55-73
- Eding, H., Meuwissen T. H. E., 2001. Markers based estimates of between and within population kinships for the conservation of genetic diversity, *J. Anim. Breed. Genet.* 118. 141-159
- Ellegren, H., 1995. Mutation rates at porcine microsatellite loci, *Mamm. Genome*, 6(5), 76-77
- Epstein, J., Bichard M., 1984. in *Evolution of Domesticated Animals*, I. L. Mason, Ed: Longman, New York, pp. 145-162
- Euler, F. V., in e-mail conference hosted by Swedish Scientific Council on Biodiversity, 1999. <http://www.gencat.es/mediamb/bioind/econfsum.htm>
- Gowdy, J. M., 1997. "The value of the Biodiversity: Markets, Society and Ecosystems". *Land Economics*, 73, 25-41
- Kimura, M., Crow, J. F., 1964. The number of alleles that can be maintained in a finite population. *Genetics*, 49, 725-738
- Krohne, D. T., 2001. *General Ecology*. 2nd Edition. Wabash College, ISBN 0-534-37528-6
- Larson G., Dobney K., Albarella U., Fang M., Matisoo-Smith E., Robins J., Lowdens S., Finlayson H., Brand T., Willerslev E., Rowley-Conwy P., Andersson L., Cooper A., 2005. Worldwide phylogeography of wild board reveals multiple centers of pig domestication. *Science*, 307
- Laval, G., SanCristobal M., Chevalet C., 2002. Measuring genetic distances between breeds : use of some distances in various short term evolution models. *Genet. Sel. Evol.*, 34, 481-507

- Laval, G., 2001. Eléments de choix des marqueurs et des méthodes dans l'analyse de diversité génétique intra spécifique : Cas des races animales domestiques. Thèse, INRA, département de Genetique Animale, Laboratoire de Génétique cellulaire
- Mainwaring, L., 2001. Biodiversity, Biocomplexity, and the Economics of genetics dissimilarity. *Land Economics*, 77, 79-93
- Malécot, G., 1946. La consanguinité dans une population limitée, *C. R. Acad. Paris*, 222, 841-843
- Malécot, G., 1948. *Les Mathématiques de l'Hérédité*. Masson et Cie, Paris, France
- Nei, M., 1971. Interspecific gene differences and evolutionary time estimated from electrophoretic data on protein identity. *The American Naturalist*, 106, 283-292
- Nei, M., 1972. Genetic distance between populations. *Am. Naturalist*, 106, 283-292
- Nei, M., 1975. *Molecular population genetics and evolution*. North-Holland Publishing Company, Amsterdam Oxford
- Nei, M., 1987. *Molecular evolutionary genetics*. Columbia University Press. ISBN 0-231-06320-2
- Nikolic, N., Park, Y. S., SanCristobal, M., Lek, S., Chevalet, C., 2009. What do artificial neural networks tell us about the genetic structure of populations? The example of European pig populations. *Genetics Research*, 91, 121-132.
- Nunney, L., 2000. The limits to knowledge in conservation genetics, The value of effective population size. *Evolutionary Biology*, 32, 179-194
- Ollivier, L., Foulley, J-L. 2005. Aggregate diversity: New approach combining within- and between-breed genetic diversity. *Livestock Production Science*. In press
- Reynolds, J., Weir, B.S., Cockerham, C.C., 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105, 767-779
- Saitou, N., Nei, N., 1987, The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology Evolution* 4, 406-425
- SanCristobal, M., Chevalet, C., Haley, C. S., Joosten, R., Rattink, A.P., Harlizius, B., Groenen, M.A.M., Amigues, Y., Boscher, M-Y., Russell, G., Law, A., Davoli, R., Russo, V., Désautés, C., Alderson, L., Fimland, E., Bagga, M., Delgado, J.V., Vegapla, J.L., Martinez, A.M., Ramos, M., Glodek, P., Meyer, J.N., Gandini, G.C., Matassino, D., Plastow, G.S., Siggens, K., Archibald, A.L., Milan, D., Hammond, K., Cardellino, R., Foulley, J-L., Ollivier, L., 2005. Genetic diversity within and between European pig breeds using microsatellite markers. *Animal Genetics*
- Towsend, C. R., Harper, J. L., Began, M., 2001. *Essentials of ecology*, Blackwell Science

Weitzman, M. L., 1992. "On diversity". *Quarterly Journal of economics*, 107, 363-405

Weitzman, M. L., 1993. "What to preserve? An application of diversity theory to crane conservation". *Quarterly Journal of economics*, 108, 157-183

Wright, S., 1943. Isolation by distance. *Genetics*, 31, 114-138

## Annexe A

### Expression de la forme quadratique

$c' \tilde{f} c$  est le produit d'un vecteur ligne, de la matrice  $\tilde{f}$  et d'un vecteur colonne :

Vecteur ligne \* Matrice \* Vecteur colonne

$$\left[ \begin{array}{cccc} c_1 & \dots & \dots & c_i \end{array} \right] \begin{pmatrix} f_{1,1} + \dots + f_{1,j} \\ \dots \\ \dots \\ f_{i,1} + \dots + f_{i,j} \end{pmatrix} \begin{pmatrix} c_1 \\ \dots \\ \dots \\ c_j \end{pmatrix}$$

$$= \left[ \begin{array}{cccc} c_1 & \dots & \dots & c_i \end{array} \right] \begin{pmatrix} \sum_j f_{1,j} c_j \\ \dots \\ \dots \\ \sum_j f_{i,j} c_j \end{pmatrix}$$

$$= c_1 \sum_j f_{1,j} c_j + \dots + c_i \sum_j f_{i,j} c_j$$

$$= \sum_i c_i \sum_j f_{i,j} c_j$$

$$= \sum_i \sum_j c_i f_{i,j} c_j$$

$$= \sum_i c_i (\sum_{j \neq i} f_{i,j} c_j + f_{i,i} c_i)$$

$$= \sum_i c_i^2 f_{i,i} + \sum_{j \neq i} c_i c_j f_{i,j}$$

Cette dernière équation est la forme développée de l'expression  $c' \tilde{f} c$

## Annexe B

### Calcul de la fonction de diversité $V(S)$ et des contributions par populations selon la méthode de Weitzman (1992-1993)

Prenons un exemple de trois populations avec pour distance génétique la distance standard de Nei (1972).

Populations	1	2	3	
1	0	0.7	0.2	} Distances standard de Nei
2	0.7	0	0.4	
3	0.2	0.4	0	

Calcul de la fonction de diversité ( $V(S)$ )

$$V(S) = V(1,2,3) = d(1,3) + \max[V(2,3), V(1,2)]$$

$$V(S) = 0,2 + 0,7 = 0,9$$

Calcul des contributions  $C1$  pour la population 1,  $C2$  pour la population 2 et  $C3$  pour la population 3. Sachant  $V(S/i)$  la fonction de diversité sans la population  $i$ .

$$C1 = 1 - V(S/1)/V(S)$$

$$C1 = 1 - (0,4/0,9) = 0,55$$

$$C2 = 1 - V(S/2)/V(S)$$

$$C2 = 1 - (0,2/0,9) = 0,77$$

$$C3 = 1 - V(S/3)/V(S)$$

$$C3 = 1 - (0,7/0,9) = 0,22$$

$$C1 = 1 - (0,4/0,9) = 0,55$$

Selon ces résultats la population qui devrait être conservée primordialement parmi les trois est la population 2.

## Annexe C

### Le coefficient de parenté de Malécot (1948) ( $J_{xy}$ ) est égal au $f_{ij}$ de Caballero et Toro (2002)

Prenons un exemple de deux populations ( $X$  et  $Y$ ) composées de trois individus chacune. Faisons les calculs pour un locus ( $Li$ ) à trois allèles ( $all1, all2$  et  $all3$ ).

	Pop. X		Pop.y	
	$Li$		$Li$	
<b>Ind. 1</b>	<i>all1</i>	<i>all2</i>	<i>all2</i>	<i>all2</i>
<b>Ind. 2</b>	<i>all2</i>	<i>all1</i>	<i>all1</i>	<i>all1</i>
<b>Ind. 3</b>	<i>all3</i>	<i>all3</i>	<i>all2</i>	<i>all1</i>

#### Valeur du coefficient de Malécot :

$$J_{xy} = \sum_i X_i Y_i$$

Fréquences alléliques  $X_i$  et  $Y_i$

$$X_1 = 2/6$$

$$X_2 = 2/6$$

$$X_3 = 2/6$$

$$Y_1 = 3/6$$

$$Y_2 = 3/6$$

$$Y_3 = 0$$

$$J_{xy} = X_1 Y_1 + X_2 Y_2 + X_3 Y_3 \cong 0,33$$

#### Valeur de la covariance entre $i$ et $j$ :

$$f_{ij} = \sum_{all} \tilde{p}_{x_i} \tilde{p}_{y_i}$$

$$\hat{p}_{k,i} = \frac{\sum_{l=1}^{N_i} P_{all,l,i}}{N_i}$$

sachant que  $\hat{p}_{k,i}$  est une fréquence génotypique prenant pour valeur 1 lorsque l'individu  $l$ , de la population  $i$ , est homozygote pour l'allèle  $all$  ; la valeur 0,5 si l'individu est hétérozygote pour l'allèle  $all$  et une valeur nulle si l'individu est homozygote pour un autre allèle.



$$\tilde{p}x_1 = \frac{(1/2 + 1/2)}{3} = 1/3$$

$$\tilde{p}x_2 = 1/3$$

$$\tilde{p}x_3 = 1/3$$

$$\tilde{p}y_1 = 1/2$$

$$\tilde{p}y_2 = 1/2$$

$$\tilde{p}y_3 = 0$$

$$F_{ij} = \tilde{p}x_1 \tilde{p}y_1 + \tilde{p}x_2 \tilde{p}y_2 + \tilde{p}x_3 \tilde{p}y_3 \cong 0,33$$

**Conclusion :**

Le coefficient de parenté de Malécot est égal au *fij* de Caballero et Toro

$$f_{ij} = J_{xy}$$

## Annexe D

Contributions d'une simulation, obtenues par l'approche de Caballero et Toro (2002), et du taux d'hétérozygotie par populations. Les populations ont la même taille efficace ( $N_e=30$ ) et sont sous dérive génétique. Trois types de simulations sont représentés avec un nombre de populations variant de 8, 14 et 24, Pour chaque simulation 50 locus partant de 400 allèles fondateurs.

8 populations			14 populations			24 populations		
Population	Hétérozygotie (%)	Contribution (%)	Population	Hétérozygotie (%)	Contribution (%)	Population	Hétérozygotie (%)	Contribution (%)
1	20,87	13,06	1	17,73	4,71	1	21,87	<b>0,00</b>
2	18,00	7,11	2	26,93	10,42	2	26,40	<b>0,00</b>
3	30,80	14,75	3	14,67	8,45	3	20,33	<b>0,00</b>
4	35,87	18,43	4	18,13	8,67	4	22,07	4,21
5	27,40	14,18	5	21,07	<b>0,00</b>	5	26,67	6,10
6	20,33	6,24	6	19,87	12,20	6	20,87	1,93
7	17,87	13,64	7	21,47	6,45	7	20,33	7,99
8	23,13	12,59	8	24,33	6,65	8	20,20	6,93
			9	26,87	4,45	9	29,87	5,77
			10	32,27	8,78	10	25,87	4,94
			11	22,87	10,40	11	20,47	4,10
			12	22,40	5,00	12	24,80	9,77
			13	29,20	8,37	13	23,13	3,52
			14	19,07	5,44	14	27,67	2,03
						15	26,67	5,13
						16	27,87	3,65
						17	26,80	7,22
						18	14,87	4,82
						19	22,47	5,19
						20	20,07	<b>0,00</b>
						21	31,27	8,20
						22	24,80	1,92
						23	22,20	7,17
						24	20,73	<b>0,00</b>
Coefficient de corrélation		0,75	Coefficient de corrélation		0,13	Coefficient de corrélation		0,25

## Annexe E

Contributions moyenne en pourcentage, obtenues par l'approche de Caballero et Toro (2002), des 1000 simulations avec le taux d'hétérozygotie moyen et l'écart type moyen. Les populations sont des populations sous dérive génétique avec un taille efficace ( $N_e$ ) égal à 30 individus. Simulations sous 50 locus de 400 allèles fondateurs.

3 populations				16 populations				24 populations			
Pop	Hétérozygotie	Contributions	Ecart-type	Pop	Hétérozygotie	Contributions	Ecart-type	Pop	Hétérozygotie	Contributions	Ecart-type
	moyenne (%)	moyenne (%)	moyen		moyenne (%)	moyenne (%)	moyen		moyenne (%)	moyenne (%)	moyen
1	22,50	33,39	3,60	1	22,40	6,30	3,48	1	22,53	4,28	3,24
2	22,24	33,06	3,47	2	22,35	6,33	3,64	2	22,30	4,20	3,26
3	22,57	33,54	3,45	3	22,33	6,17	3,52	3	22,38	4,01	3,15
				4	22,33	6,17	3,35	4	22,22	4,11	3,12
				5	22,13	6,13	3,49	5	22,38	4,32	3,29
				6	22,61	6,46	3,50	6	22,19	4,10	3,18
				7	22,28	6,31	3,46	7	22,48	4,11	3,25
				8	22,52	6,25	3,52	8	22,26	4,33	3,14
				9	22,48	6,30	3,44	9	22,36	4,10	3,16
				10	22,28	6,19	3,56	10	22,41	4,07	3,03
				11	22,30	6,08	3,58	11	22,26	3,99	3,10
				12	22,44	6,24	3,56	12	22,78	4,28	3,22
				13	22,20	6,22	3,46	13	22,11	4,23	3,26
				14	22,39	6,39	3,50	14	22,07	4,10	3,18
				15	22,60	6,26	3,61	15	22,37	3,94	3,17
				16	22,44	6,20	3,78	16	22,30	4,23	3,22
								17	22,46	4,21	3,27
								18	22,34	4,18	3,23
								19	22,47	4,41	3,25
								20	22,38	4,15	3,19
								21	22,48	4,12	3,20
								22	22,59	4,42	3,26
								23	22,59	4,12	3,21
								24	22,34	3,99	3,87

## Annexe F

Contributions obtenues par l'approche de Weitzman (1992-1993) avec des populations sous dérive génétique et une taille efficace faible ( $N_e=30$  individus). Simulations avec 50 locus de 400 allèles fondateurs.

8 populations			14 populations			24 populations		
Populations	Hétérozygotie (%)	Contribution (%)	Populations	Hétérozygotie (%)	Contribution (%)	Populations	Hétérozygotie (%)	Contribution (%)
1	20,87	13,08	1	17,73	5,60	1	21,87	3,47
2	18,00	11,72	2	26,93	7,01	2	26,40	2,87
3	30,80	10,97	3	14,67	9,06	3	20,33	3,22
4	35,87	12,85	4	18,13	7,50	4	22,07	3,97
5	27,40	13,80	5	21,07	5,48	5	26,67	4,72
6	20,33	11,03	6	19,87	8,27	6	20,87	3,77
7	17,87	17,27	7	21,47	7,57	7	20,33	5,98
8	23,13	12,43	8	24,33	5,45	8	20,20	4,83
			9	26,87	6,84	9	29,87	3,96
			10	32,27	5,11	10	25,87	4,44
			11	22,87	8,23	11	20,47	4,39
			12	22,40	7,08	12	24,80	4,66
			13	29,20	6,01	13	23,13	3,14
			14	19,07	7,89	14	27,67	3,52
						15	26,67	3,24
Coefficient de corrélation		-0,26	Coefficient de corrélation		-0,57	16	27,87	3,76
						17	26,80	4,46
						18	14,87	4,35
						19	22,47	4,39
						20	20,07	3,70
						21	31,27	4,31
						22	24,80	3,23
						23	22,20	3,97
						24	20,73	3,51
						Coefficient de corrélation		-0,14

## Annexe G

Contributions obtenues par l'approche de Caballero et Toro (2002) et taux d'hétérozygotie de populations à l'équilibre dérive et mutation lorsque le nombre de populations varie. Pour chacune des simulations 50 locus de 400 allèles fondateurs.

8 populations			16 populations			24 populations				
Populations	Hétérozygotie	Contributions	Populations	Hétérozygotie	Contributions	Populations	Hétérozygotie	Contributions		
1	71,99	12,48	1	68,33	6,46	1	75,69	5,54		
2	67,16	11,88	2	71,36	7,03	2	71,76	4,51		
3	73,15	13,69	3	69,19	5,88	3	68,38	3,54		
4	71,95	13,70	4	68,28	6,15	4	72,17	4,99		
5	69,53	12,38	5	68,62	6,14	5	71,31	4,43		
6	66,43	11,79	6	66,00	5,40	6	65,22	3,56		
7	67,71	11,65	7	65,21	5,43	7	64,11	3,73		
8	69,57	12,44	8	71,63	6,69	8	69,84	3,80		
			9	69,71	6,62	9	71,03	4,31		
Coefficient de corrélation			0,89	10	68,81	5,87	10	70,39	4,31	
			11	66,27	6,04	11	69,22	4,38		
			12	70,71	6,72	12	70,39	4,62		
			13	73,49	6,27	13	72,72	4,88		
			14	67,66	6,00	14	69,61	3,93		
			15	71,11	6,24	15	65,13	3,57		
			16	69,93	7,06	16	70,36	3,99		
			Coefficient de corrélation			0,69	17	69,79	3,93	
							18	68,69	4,17	
							19	62,11	3,51	
							20	69,01	3,96	
							21	68,54	4,30	
							22	67,87	3,78	
							23	67,50	4,12	
							24	67,85	4,14	
							Coefficient de corrélation			0,84

## Annexe H

Contributions moyenne en pourcentage, obtenues par l'approche de Caballero et Toro (2002), de 1000 simulations avec le taux d'hétérozygotie moyen en pourcentage et l'écart type moyen. Les populations sont des populations sous équilibre dérive et mutation avec une taille efficace ( $N_e$ ) égale à 300 individus. Simulations sous 50 locus de 400 allèles fondateurs.

4 populations				8 populations				24 populations			
Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen	Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen	Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen
1	69,00	24,97	1,60	1	69,02	12,51	0,91	1	68,94	4,16	0,45
2	69,13	25,04	1,59	2	69,06	12,52	0,93	2	69,04	4,15	0,46
3	68,94	24,95	1,59	3	68,90	12,47	0,92	3	69,10	4,20	0,44
4	69,10	25,03	1,60	4	69,06	12,54	0,93	4	68,98	4,17	0,44
				5	69,03	12,51	0,95	5	68,99	4,17	0,43
				6	68,94	12,47	0,94	6	69,01	4,16	0,45
				7	69,00	12,50	0,94	7	69,10	4,17	0,45
				8	69,01	12,47	0,92	8	68,88	4,18	0,45
								9	69,09	4,17	0,45
								10	69,06	4,18	0,44
								11	68,97	4,14	0,44
								12	69,13	4,18	0,45
								13	69,10	4,18	0,45
								14	68,95	4,14	0,45
								15	69,05	4,17	0,46
								16	69,08	4,19	0,46
								17	69,13	4,19	0,45
								18	69,05	4,19	0,47
								19	68,88	4,15	0,47
								20	69,02	4,17	0,43
								21	69,04	4,16	0,45
								22	68,89	4,14	0,45
								23	68,95	4,14	0,44
								24	68,99	4,17	0,46

# Annexe I

Contributions en pourcentage, obtenues par l'approche de Weitzman (1992-1993), et taux d'hétérozygotie des populations à l'équilibre dérive et mutation dont le nombre de populations varie. Pour chacune des simulations 50 locus de 400 allèles fondateurs.

<b>8 populations</b>			<b>16 populations</b>			<b>24 populations</b>		
Populations	Hétérozygotie (%)	Contributions (%)	Populations	Hétérozygotie (%)	Contributions (%)	Populations	Hétérozygotie (%)	Contributions (%)
1	71,99	13,73	1	68,33	6,70	1	75,69	3,67
2	67,16	15,42	2	71,36	6,14	2	71,76	4,25
3	73,15	13,73	3	69,19	6,57	3	68,38	4,42
4	71,95	13,22	4	68,28	7,02	4	72,17	3,83
5	69,53	13,95	5	68,62	6,69	5	71,31	3,98
6	66,43	15,67	6	66,00	7,16	6	65,22	4,88
7	67,71	14,48	7	65,21	7,71	7	64,11	5,14
8	69,57	14,33	8	71,63	6,31	8	69,84	4,13
			9	69,71	6,54	9	71,03	3,97
Coefficient de corrélation		-0,90	10	68,81	6,80	10	70,39	4,08
			11	66,27	7,48	11	69,22	4,32
			12	70,71	6,10	12	70,39	4,18
			13	73,49	6,01	13	72,72	3,87
			14	67,66	6,85	14	69,61	4,26
			15	71,11	6,23	15	65,13	4,98
			16	69,93	6,33	16	70,36	4,25
						17	69,79	4,37
			Coefficient de corrélation		-0,95	18	68,69	4,39
						19	62,11	5,25
						20	69,01	4,28
						21	68,54	4,20
						22	67,87	4,47
						23	67,50	4,58
						24	67,85	4,55
						Coefficient de corrélation		-0,96

## Annexe J

Contributions moyenne en pourcentage, obtenues par l'approche de Weitzman(1992-1993), des 1000 simulations avec le taux d'hétérozygotie moyen et l'écart type moyen. Les populations sont des populations sous équilibre dérive et mutation avec une taille efficace ( $N_e$ ) égale à 300 individus. Simulations sous 50 locus de 400 allèles fondateurs.

3 populations				16 populations				24 populations			
Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen	Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen	Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen
1	69,00	30,80	1,77	1	69,02	30,77	1,98	1	68,94	30,74	2,14
2	69,13	30,73	1,78	2	69,06	30,77	2,01	2	69,04	30,66	2,09
3	68,94	30,82	1,76	3	68,90	30,89	2,04	3	69,10	30,59	2,16
4	69,10	30,72	1,77	4	69,06	30,69	2,06	4	68,98	30,72	2,13
				5	69,03	30,76	2,03	5	68,99	30,70	2,14
				6	68,94	30,83	2,09	6	69,01	30,72	2,14
				7	69,00	30,75	1,98	7	69,10	30,58	2,12
				8	69,01	30,78	2,04	8	68,88	30,73	2,25
								9	69,09	30,63	2,12
								10	69,06	30,69	2,15
								11	68,97	30,69	2,10
								12	69,13	30,62	2,11
								13	69,10	30,61	2,16
								14	68,95	30,74	2,15
								15	69,05	30,62	2,21
								16	69,08	30,57	2,19
								17	69,13	30,60	2,13
								18	69,05	30,65	2,22
								19	68,88	30,77	2,20
								20	69,02	30,65	2,14
								21	69,04	30,64	2,18
								22	68,89	30,82	2,23
								23	68,95	30,69	2,16
								24	68,99	30,71	2,18



## Annexe K

Contributions moyenne, obtenues par l'approche de Weitzman(1992-1993), par populations des 1000 simulations avec le taux d'hétérozygotie moyen et l'écart type moyen. Les populations sont des populations sous dérive génétique avec une taille efficace ( $N_e$ ) égale à 30 individus. Simulations sous 50 locus de 400 allèles fondateurs.

3 populations				16 populations				24 populations			
Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen	Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen	Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen
1	22,50	50,00	6,46	1	22,40	6,20	1,16	1	22,53	1,01	0,21
2	22,24	49,71	6,45	2	22,35	6,28	1,26	2	22,30	1,01	0,21
3	22,57	50,42	6,37	3	22,33	6,21	1,20	3	22,38	1,00	0,21
				4	22,33	6,15	1,07	4	22,22	1,01	0,21
				5	22,13	6,22	1,16	5	22,38	1,02	0,22
				6	22,61	6,29	1,23	6	22,19	1,01	0,20
				7	22,28	6,25	1,21	7	22,48	1,00	0,20
				8	22,52	6,25	1,21	8	22,26	1,01	0,20
				9	22,48	6,22	1,19	9	22,36	1,00	0,21
				10	22,28	6,22	1,19	10	22,41	1,00	0,20
				11	22,30	6,15	1,16	11	22,26	1,00	0,21
				12	22,44	6,19	1,18	12	22,78	1,01	0,21
				13	22,20	6,20	1,17	13	22,11	1,02	0,22
				14	22,39	6,27	1,21	14	22,07	1,01	0,20
				15	22,60	6,21	1,25	15	22,37	0,99	0,20
				16	22,44	6,19	1,15	16	22,30	1,01	0,22
								17	22,46	1,01	0,21
								18	22,34	1,01	0,20
								19	22,47	1,02	0,21
								20	22,38	1,00	0,20
								21	22,48	1,00	0,20
								22	22,59	1,01	0,20
								23	22,59	1,00	0,19
								24	22,34	1,02	0,23

8 populations			
Pop	Hétérozygotie moyen (%)	Contributions moyenne (%)	Ecart-type moyen
1	22,36	13,65	2,51
2	22,47	13,50	2,57
3	22,56	13,57	2,48
4	22,48	13,54	2,55
5	22,60	13,49	2,54
6	22,39	13,58	2,61
7	22,22	13,48	2,57
8	22,44	13,56	2,59

