



HAL
open science

Création d'un observatoire de la variabilité génétique des ruminants et des équidés

Coralie Danchin-Burge

► **To cite this version:**

Coralie Danchin-Burge. Création d'un observatoire de la variabilité génétique des ruminants et des équidés. Innovations Agronomiques, 2013, 29, pp.61-74. 10.17180/8t1q-k681 . hal-04531001

HAL Id: hal-04531001

<https://hal.inrae.fr/hal-04531001v1>

Submitted on 3 Apr 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Création d'un observatoire de la variabilité génétique des ruminants et des équidés

Danchin-Burge C. ¹

¹ Institut de l'Élevage, Département Génétique et Phénotypes, 149 rue de Bercy, 75595 Paris Cedex 12

Correspondance : coralie.danchin@idele.fr

Résumé

Dans les espèces d'élevage, les programmes de sélection sont de plus en plus performants, et l'essor de nouveaux outils tels que la sélection génomique est de nature à accélérer nettement l'évolution de ces populations. Or, il est indispensable de préserver la variabilité génétique de ces populations, d'une part parce que le progrès génétique est corrélé à ce facteur, et d'autre part parce que l'augmentation brutale de la consanguinité dans les populations peut les impacter très défavorablement. Il est donc nécessaire d'assurer un suivi régulier de la variabilité génétique, pour garantir la durabilité des choix de sélection de demain, et ce aussi bien pour les populations à faible effectif que celles en sélection. Pour cela, deux types de données permettent le calcul d'indicateurs de variabilité génétique : les généalogies et les données de génotypages. Ces indicateurs permettent, par exemple, d'évaluer l'impact d'un ancêtre majeur sur la population actuelle. L'objectif du projet VARUME (VARIabilité génétique des RUMinants et des Equidés) est la mise en place d'un observatoire de la variabilité génétique des ruminants et équidés, afin de produire de façon régulière des indicateurs permettant d'apprécier la variabilité génétique des races, sur des bases communes reconnues par l'ensemble des gestionnaires.

Mots-clés : variabilité génétique, observatoire, ruminants, équidés, généalogies, marqueurs moléculaires

Abstract: Setting up a genetic variability observatory for ruminants and equids species

In livestock species, selection programs are getting more and more efficient. Meanwhile, the development of new tools such as genomic selection will probably fasten how breeds are evolving. However, preserving the breeds' genetic variability is a necessity. First of all, genetic progress is correlated with this factor. Also, a sharp increase in inbreeding in a breed might have a very negative impact. Therefore, in order to achieve the sustainability of our selection choices for the future, the breeds' genetic variability requires an accurate management. To do so, two type or data can be used to calculate variability genetic indicators: pedigrees and genotype data. These indicators can be used to monitor the impact of a main ancestor in a breed for instance. The VARUME (Genetic VARIability of Ruminants and Equids) project goal is to set up a genetic variability observatory for the Ruminants and Equids species. It will generate indicators that can assess breed genetic variability on a regular basis, by using a common method that will be acknowledged and used by all breed managers.

Keywords: genetic variability, observatory, ruminants, equids pedigrees, molecular markers

Introduction

La protection de la biodiversité est l'un des enjeux majeurs actuels, comme le montre le choix de l'Organisation des Nations Unies de proclamer 2010 « Année internationale de la biodiversité ». Cependant, dans les actions mises en place, la protection de la biodiversité créée par l'homme, à savoir, pour les espèces animales, les populations animales sélectionnées, encore appelée Ressources Génétiques Animales (RGA), est souvent oubliée. Pourtant les RGA font bien partie intégrante de la Convention sur la Diversité Biologique que la France a ratifiée en 1992.

En France, les programmes de sélection des espèces de ruminants et d'équidés se sont révélés très efficaces et ont apporté une contribution majeure au développement des filières. Organisés localement, à une échelle nationale voire mondiale, ces programmes sont caractérisés par une concentration des efforts de sélection sur une élite réduite, sur le plan numérique, de reproducteurs, conduisant à des tailles génétiques efficaces très réduites. Cette situation est particulièrement accentuée dans les filières où l'Insémination Animale (IA) est utilisée majoritairement et dans les espèces où elle autorise une très large diffusion d'un reproducteur donné. Ainsi les populations animales sélectionnées ont connu des pertes de variabilité génétique plus ou moins importantes. Les conséquences peuvent être des effets délétères sur la santé des animaux (dépression de consanguinité) ainsi qu'une réduction du pool d'allèles que constituent ces populations, et donc des possibilités limitées de répondre aux besoins futurs de sélection et d'adaptation.

La volonté de continuer à effectuer des progrès génétiques, tout en se donnant la possibilité de réorienter les objectifs de sélection, nécessite de maintenir la variabilité génétique au sein de chaque race. Parallèlement, la survie des races à petits effectifs requiert une surveillance accrue de l'évolution de leur variabilité génétique.

Plusieurs types d'information et différentes méthodes permettent d'apprécier la variabilité génétique au sein d'une population. On trouve dans la littérature de nombreuses études de variabilité génétique des races de ruminants à partir de données généalogiques, que ce soit pour les ovins (Selavaggi *et al.*, 2010 ; Danchin-Burge *et al.*, 2010), les bovins (Avon et Colleau, 2006 ; Koenig and Simianer 2006 ; Mc Parland *et al.*, 2007 ; Danchin-Burge *et al.*, 2011) ou les caprins (Piacère *et al.*, 2004 ; Danchin-Burge *et al.*, 2012). Pour les équidés, des études de variabilité génétique à base de données généalogiques ont également été réalisées pour plusieurs races de sport (Valera *et al.*, 2005 ; Bokor *et al.*, 2013), ou encore des races de chevaux de trait et d'ânes (Poncet *et al.*, 2006 ; Rizzi *et al.*, 2011).

Toutefois, pour la France, ces études ont été réalisées au cas par cas, par des intervenants variés, généralement peu valorisées et rarement réfléchies de façon globale. Le détachement à l'INRA d'un ingénieur de l'Institut de l'Élevage (C. Danchin-Burge) a permis pour la première fois que ces bilans soient réalisés de façon exhaustive pour l'ensemble des populations de ruminants (Institut de l'Élevage, 2009, 2010). Pour les équidés, une réflexion équivalente était en cours à l'Institut Français du Cheval et de l'Équitation (IFCE).

Par ailleurs, les progrès récents accomplis dans la connaissance du génome avec la possibilité de caractériser plus finement des animaux grâce à une forte densité de marqueurs moléculaires ont permis de créer une base très fournie d'information potentielle sur la variabilité génétique des individus et des populations. Or, ces marqueurs sont peu exploités en ce sens.

Dans cet article, nous rappellerons d'abord pourquoi la conservation de la variabilité génétique est cruciale pour la sélection animale, puis quelles sont les sources d'information et les indicateurs de variabilité génétique disponibles et nous présenterons enfin le projet VARUME : Observatoire de la VARIabilité génétique des RUMinants et des Equidés, dont l'objectif est la création d'un observatoire de la variabilité génétique, avec ces deux volets, observatoire généalogique et observatoire moléculaire.

1. Pourquoi conserver la variabilité génétique ?

1.1. La France, terre de biodiversité et de sélection d'élevage

Les programmes de sélection des espèces de ruminants existent en France depuis les années 1960 et ont été encouragés notamment par la loi sur l'élevage de 1966. Basés sur le repérage, la création, l'utilisation et la diffusion des meilleurs reproducteurs, ils ont parfois conduit à une réduction importante de la diversité génétique de certaines races (Danchin-Burge *et al.*, 2011; Danchin-Burge *et al.*, 2012). Les populations ainsi sélectionnées répondent aux exigences actuelles de l'élevage mais qu'en sera-t-il à l'avenir ? Il est stratégique de maintenir une variabilité génétique de ces populations afin qu'elles puissent s'adapter aux évolutions que connaîtront les objectifs de sélection dans les années à venir.

Pour les races à faibles effectifs, la France peut se féliciter d'être l'un des pays d'Europe qui a su maintenir l'un des plus grands nombres de races locales notamment dans les espèces de ruminants. Les programmes de conservation lancés à partir de la fin des années 70 ont permis à ces races aujourd'hui dites à petits ou très petits effectifs de se maintenir et même de se développer. Par exemple, pour les bovins à très faibles effectifs (15 races avec moins de 1 000 vaches reproductrices par race), le nombre de vaches a été multiplié par 8 en 30 ans (données Institut de l'Élevage). La richesse du réservoir génétique constitué par ces « petites » races est donc très grande en terme de variabilité génétique. Néanmoins, le maintien de ces populations dépend aussi de la surveillance de l'état de leur variabilité génétique et celle-ci n'est pas organisée en routine. Ces problématiques se rencontrent également chez les équidés. Ainsi, parmi les races à petits effectifs d'origine française, on ne compte pas moins de 9 races de chevaux de trait, 2 races de poneys et 4 races de selle, ainsi que 7 races d'âne (Danvy et Heydeman, 2005). Parmi les races de sport, le selle français, emblème de la qualité de notre élevage, a connu une forte introgression de races extérieures dans les années 2000, et l'impact de cet apport n'a pas encore été quantifié.

1.2. La variabilité génétique, source de sélection

Mais en fait, est-ce que la variabilité génétique est bien nécessaire, et en particulier aux populations en sélection ? La réponse est clairement oui, pour les raisons exposées ci-dessous.

D'après la théorie de la génétique quantitative, pour un caractère donné, le progrès génétique (ΔG) attendu sous l'effet de la sélection est fonction des paramètres suivants :

$$E(\Delta G) = \frac{i\rho\sigma_A}{t}$$

Où i est l'intensité de sélection ;

ρ la précision du critère de sélection ;

σ_A , l'écart-type génétique additif du caractère sélectionné ;

T , l'intervalle de génération

Le progrès génétique dépend donc strictement de la variabilité génétique du caractère sélectionné. Gérer la variabilité génétique d'une population en sélection, c'est préserver son potentiel de progrès génétiques (à la fois pour les caractères sélectionnés aujourd'hui et pour ceux que l'on souhaiterait sélectionner demain). Pour les races menacées, en dehors des aspects patrimoniaux, du point de vue génétique, il s'agit, d'une part, de maintenir le potentiel d'adaptation de ces populations et, d'autre part, de conserver des ressources génétiques pour d'éventuels besoins futurs.

La gestion par race des animaux d'élevage conduit à l'accouplement d'animaux semblables, en se référant en général à un standard, du moins en France et dans la plupart des pays industrialisés. Or, l'effectif de chaque race étant limité, on aboutit forcément à l'accouplement entre animaux apparentés,

et donc à la procréation d'animaux dits consanguins dans la population. Ce phénomène est plus ou moins accentué en fonction du nombre de reproducteurs utilisés dans la race, du déséquilibre du sexe-ratio, et des choix réalisés pour les accouplements. Dans une population fermée, on assiste donc inéluctablement à une diminution de la variabilité génétique, sous les effets de la dérive génétique et de l'augmentation de la consanguinité, qui sont les deux facettes d'un même phénomène :

- La dérive génétique est due au fait que, d'une génération à l'autre le choix d'un nombre limité de reproducteurs représente un échantillonnage des gènes. Cela provoque des fluctuations aléatoires des fréquences alléliques, avec d'autant plus d'ampleur que le nombre de reproducteurs est faible. A plus ou moins long terme, en un locus donné, on aboutit à la fixation d'un allèle et à la perte de tous les autres. Du fait que les fréquences alléliques vont vers les extrêmes, la proportion d'animaux homozygotes augmente à un locus donné.
- L'augmentation de la consanguinité est due à l'instauration progressive d'une parenté entre reproducteurs au sein de la population. Cet accroissement de la consanguinité s'accompagne d'une augmentation de la proportion d'homozygotes.

L'accroissement de la consanguinité et l'érosion de la variabilité génétique dépendent de plusieurs paramètres, dont les effectifs de reproducteurs des deux sexes et la variabilité de leurs tailles de descendance. On voit là clairement pourquoi les races à petits effectifs sont particulièrement concernées par la nécessité de la prise en compte de la variabilité génétique dans leur gestion. Les races en sélection sont également concernées, à plusieurs titres. Le choix de l'IA comme moyen privilégié de diffusion, en particulier pour les filières laitières, et surtout chez les bovins, implique une forte diminution du nombre de reproducteurs mâles, une augmentation de la taille moyenne de descendance de ces derniers et la possibilité d'un accroissement substantiel de la variance des tailles de descendance entre mâles. Enfin, le choix du BLUP (Best Linear Unbiased Predictor) appliqué à un modèle animal comme méthode d'évaluation des caractères sélectionnés augmente la probabilité de sélection d'animaux apparentés (Verrier *et al.*, 1993).

On appelle dépression de consanguinité les effets délétères de la consanguinité sur la viabilité d'une race. Elle peut être définie comme la réduction de la valeur moyenne d'un caractère due à la consanguinité. Elle concerne les caractères quantitatifs de manière générale mais, d'après Falconer (1981), la dépression affecte principalement les caractères faiblement héritables (vigueur, adaptation, reproduction). Un autre effet néfaste de la consanguinité est l'accroissement de la proportion d'animaux atteints par une anomalie génétique à déterminisme simple, sous le simple effet de l'augmentation du taux d'animaux homozygotes.

2. Quelles sources d'information pour créer des indicateurs de variabilité génétique ?

Plusieurs types d'informations et différentes méthodes permettent d'apprécier la variabilité génétique au sein d'une population (Boichard *et al.*, 1997 ; Baumung *et al.*, 2003). En dehors des données simples de polymorphisme visible et mesurable (poids, taille, couleur de robe, cornage, performances zootechniques...), utilisées principalement pour des populations pour lesquelles d'autres données ne peuvent être collectées, les sources de données utilisables pour calculer des indicateurs de variabilité génétique sont les généalogies ou les polymorphismes moléculaires.

2.1. Données généalogiques

Les généalogies permettent de quantifier la variabilité génétique et son évolution dans le temps en termes de probabilités d'origine ou d'identité des gènes en un endroit quelconque du génome. Elles aussi offrent l'avantage d'être facilement disponibles, sans surcoût – car collectées à d'autres fins que la mesure de la diversité génétique - et raisonnablement fiables sur de longues durées.

2.1.1. Indicateurs de qualité des généalogies

Un moyen d'estimer la fiabilité des indicateurs de variabilité génétique est de se référer à la qualité de l'information généalogique disponible. Ce facteur est fondamental : l'hypothèse de base de ce type d'analyse est que si un animal n'a pas de parents connus (ni père, ni mère), alors il est non apparenté au reste de la population. C'est un postulat fort, qui est évidemment souvent faux, et qui l'est d'autant plus quand on prend en compte un animal né récemment. Cela veut donc dire que lorsque l'on compare deux races, il faudra toujours le faire en ayant en tête le niveau respectif de la qualité des généalogies de chacune. Pour illustrer ce concept, si une race est connue, en moyenne, sur seulement deux générations, l'analyse que l'on en fera devrait en principe montrer qu'elle est beaucoup plus variable qu'une autre race avec des effectifs équivalents mais connue, elle, sur dix générations. La réalité est évidemment souvent autre.

Pour estimer la qualité des généalogies, il faut disposer d'un moyen permettant de quantifier le degré de connaissance généalogique connu par race. Cet indicateur synthétique, appelé nombre équivalent de générations connues (Ngen), est obtenu en sommant la proportion d'ancêtres i connus sur chaque génération n . Par exemple, une race dont la population analysée a un Ngen de 5 veut dire que, en moyenne, les ancêtres d'un animal issu de la population analysée sont renseignés sur 5 générations.

Il est donc indispensable de renseigner les gestionnaires quant à la fiabilité des informations dont ils disposent. On peut produire plusieurs sortes d'indicateurs de qualité :

- Un indicateur évolutif : le niveau d'information des généalogies par rang d'ascendance ;
- Des indicateurs statiques : le nombre moyen de générations remontées et d'ancêtres connus, le maximum de générations remontées...

2.1.2. Indicateurs démographiques

Des critères démographiques simples sont des notions très parlantes pour les gestionnaires. De plus, elles permettent d'avoir des indications de niveau de variabilité génétique pour des races dont les généalogies sont mal connues, et donc pour qui les indicateurs peuvent être considérés comme assez peu fiables.

Trois types d'indicateurs peuvent être proposés :

- Des **indicateurs statiques**, calculés pour chaque population analysée : nombre d'animaux (avec un parent connu, deux parents connus), et pour les animaux de la population ayant au moins un parent connu, le nombre de pères, de grands-pères paternels, etc.
- Des **indicateurs évolutifs** : évolution de la population femelle (totale, avec 1 ou 2 parents connus) ou encore l'évolution du nombre d'élevages naisseurs, par année de naissance...
- Une **pyramide des âges** des femelles actives (avec un ou deux parents connus).

2.1.3. Indicateurs issus de la probabilité d'origine des gènes

Le principe des méthodes développées pour la probabilité des gènes consiste à analyser la **provenance** des gènes de la population analysée. L'objectif est de repérer les **ancêtres majeurs** de cette population, c'est à dire les ancêtres par qui les généalogies passent préférentiellement. Pour calculer l'influence de chaque ancêtre, on considère :

- que les fondateurs (ancêtres sans ascendants) sont la source de tous les gènes actuels (pas de prise en compte de mutations éventuelles).
- qu'un gène d'un individu X provient de son père ou de sa mère, avec une probabilité de $\frac{1}{2}$; d'un des quatre grands-parents avec une probabilité de $\frac{1}{4}$, etc.

L'espérance de la contribution d'un ancêtre au génome d'un ou plusieurs individus est appelée **probabilité d'origine des gènes**. Pour chacun des ascendants, est calculée la probabilité d'avoir transmis un gène à l'individu X de la population analysée. Puis, en sommant les probabilités de transmission des gènes, on calcule la contribution de chaque ancêtre à la population de référence. On calcule d'abord la contribution d'un ancêtre sans prendre en compte ses relations de parenté avec les autres animaux (contribution brute) puis on prend en compte les relations de parenté pour calculer sa contribution marginale. On peut estimer, à partir de ces valeurs, **un nombre efficace d'ancêtres** qui correspond au nombre d'ancêtres, qui en ayant tous des contributions parfaitement égales, engendreraient une population du même niveau de variabilité génétique que la population de référence. Un des avantages des critères issus de la probabilité d'origine des gènes est qu'ils ne sont pas trop dépendants du niveau de qualité des généalogies. En revanche, par rapport aux critères issus de la probabilité d'identité des gènes (voir paragraphe suivant), ils illustrent plutôt des évolutions passées de la variabilité génétique. Le Tableau 1 illustre un bilan des indicateurs issus de la probabilité d'origine des gènes, avec l'exemple de la race bovine Jersiaise.

Tableau 1 : Indicateurs de variabilité génétique issus de la probabilité d'origine des gènes, pour la population analysée suivante : femelles de race Jersiaise avec deux parents connus, nées entre 2008 et 2011

Critère	Résultat
Effectif de la population analysée	7 087
Nombre de fondateurs (f)	6 560
Nombre de fondateurs efficaces (Fe)	120
Nombre d'ancêtres efficaces (Ae)	52
Ratio Ae/Fe*	44 %
Contribution marginale de l'ancêtre principal	7,2 %
Nombre d'ancêtres expliquant 50 % de la variabilité génétique de la population analysée	21

* permet de détecter la présence de goulets d'étranglement importants, c'est à dire l'influence prépondérante de quelques animaux dans les généalogies.

Les indicateurs rendent compte de phénomènes différents. De manière systématique : $f > Fe > Ae$. Ainsi, dans notre exemple, les 7 087 femelles sont issues de 6 560 fondateurs, mais elles ont en réalité la même variabilité génétique qu'une autre race qui aurait seulement 52 ancêtres, avec des contributions équilibrées. Par ailleurs, seulement 21 ancêtres différents expliquent 50% de la variabilité génétique analysée. Les conséquences de la création de goulets d'étranglement peuvent être illustrées par les crises qui affectent régulièrement la filière bovine laitière les années 1990, avec l'apparition d'animaux porteurs d'anomalies à une fréquence élevée. Ces crises ont principalement touché la race Prim'Holstein, parce qu'il s'agit de la première race mondiale par ses effectifs. L'ancêtre responsable de la diffusion de la mutation de l'anomalie a pu être identifié chaque fois, et il s'agit toujours d'un ancêtre majeur.

Tableau 2 : Crises majeures dues à des anomalies génétiques en race Prim'Holstein, et ancêtre majeur responsable de la diffusion de la maladie.

Anomalie	Date de la crise	Ancêtre vecteur principal
Bovine Leucocyte Adhesion Deficiency (BLAD)	1992-1995	Osborndale Ivanhoe
Bulldog	1999-2001	Igale Masc
Complex Vertebral Malformation (CVM)	2000-2003	Bell
Brachyspina	2006-actuellement	Sweet Haven Tradition

La théorie de la probabilité d'origine des gènes est également utilisée pour tracer l'évolution de gènes étrangers (c'est à dire l'apport d'animaux de pays étrangers ou bien l'apport d'animaux d'une race exogène) au sein de la race étudiée. Pour cela, une race ou une origine est attribuée à chaque fondateur (c'est à dire un animal sans ascendants connus) et l'évolution de la transmission des gènes est suivie d'une génération à l'autre, avec l'approche de la probabilité d'origine des gènes. L'exemple de la race bovine Jersey, pour laquelle nous avons tracé l'évolution des apports américains et danois dans la race (Figure 1) permet d'illustrer cette approche.

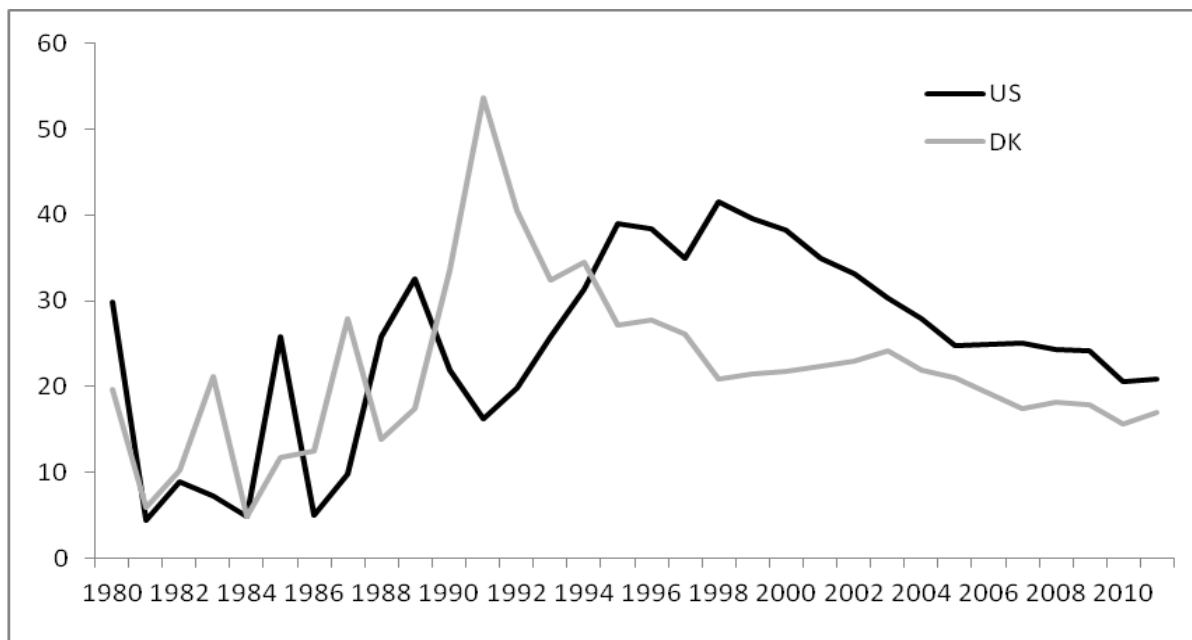


Figure 1 : Evolution des apports de gènes danois (DK) et américains (US) (en %, échelle des ordonnées) dans la race bovine Jersey.

Dans un premier temps, les gènes danois ont pris le pas sur l'influence américaine au début des années 90, pour diminuer rapidement et être substitués partiellement par les gènes américains. Depuis les années 2000, l'influence des gènes des animaux provenant de ces deux pays est en diminution, signifiant l'existence probable d'une autre origine, sans doute australe (Australie, Nouvelle-Zélande).

2.1.4. Indicateurs issus de la probabilité d'identité des gènes : parenté et consanguinité

La question sous jacente au calcul de ces indicateurs est de savoir si les allèles portés par un individu sont identiques. Deux animaux sont apparentés s'ils partagent un ancêtre commun. Un animal est consanguin si ses parents sont apparentés. Deux allèles sont identiques par descendance s'ils sont issus de la duplication du même allèle ancestral. Le coefficient de consanguinité d'un individu est la probabilité que deux allèles portés par un animal soient identiques par descendance. Il est égal au coefficient de parenté entre ses parents. Comme l'augmentation de la consanguinité a pour conséquence d'augmenter l'homogénéisation du génome de la population, le rythme d'accroissement de la consanguinité donne une bonne indication de la vitesse de perte de variabilité génétique. Un accroissement maîtrisé de la consanguinité permet d'éliminer progressivement des gènes défavorables, sans conséquences importantes sur la survie ou la reproduction des animaux. Avec une augmentation brusque, les probabilités sont plus élevées de faire apparaître des animaux porteurs de tares.

A partir des calculs individuels de consanguinité, on peut généraliser à l'ensemble d'une race. Pour interpréter les résultats, il faut avoir en tête que le pourcentage d'animaux consanguins et le niveau de consanguinité augmentent logiquement au fur et à mesure que les généalogies sont mieux connues. On comprend donc que la question à se poser n'est pas : « à partir de quel niveau de consanguinité

une race doit faire des efforts de gestion de sa variabilité » mais plutôt : « à partir de quel seuil d'accroissement de la consanguinité une race doit faire des efforts de gestion de sa variabilité ».

Pour illustrer de manière plus parlante ce que représente l'accroissement de la consanguinité pour une période donnée, et pouvoir comparer plus aisément les races entre elles, il existe un indicateur appelé effectif génétique (N_e). Un effectif génétique réalisé de 100 correspond, par exemple, à une population dont l'effectif femelle serait de taille infinie mais dont l'effectif mâle serait de 25 seulement. Pour calculer cet indicateur il existe de nombreuses méthodes différentes qui diffèrent en général selon la façon dont l'accroissement de la consanguinité est calculé. Les résultats peuvent être très distincts suivant la méthode employée, en particulier pour les races dont l'accroissement de la consanguinité est faible ou très variable d'une année sur l'autre (Leroy *et al.*, 2012). Pour le moment, même si toutes les méthodes présentent des biais, la méthode Cervantes (Cervantes *et al.*, 2011) semble la plus pertinente, car elle se base sur l'accroissement de la parenté au lieu de la consanguinité, et qu'elle est corrigée par la connaissance généalogique individuelle (Leroy *et al.*, 2012).

2.1.5. Exemple de résultats : indicateurs par filière pour les ruminants

Les indicateurs de variabilité (Figure 2) présentent un continuum avec, du côté le plus avantageux, les races bovines allaitantes, puis les petits ruminants laitiers et les petits ruminants allaitants, et enfin les races bovines laitières. Les races à petits effectifs présentent des résultats plus hétérogènes. De façon globale, les valeurs plus ou moins élevées des indicateurs de variabilité génétique par filière sont le reflet de plusieurs facteurs :

- Le taux de pénétration de l'IA pour renouveler les reproducteurs ;
- L'effectif des reproducteurs femelles ;
- L'existence ou non de systèmes de gestion de la diversité ;
- L'introgression régulière ou non de gènes étrangers.

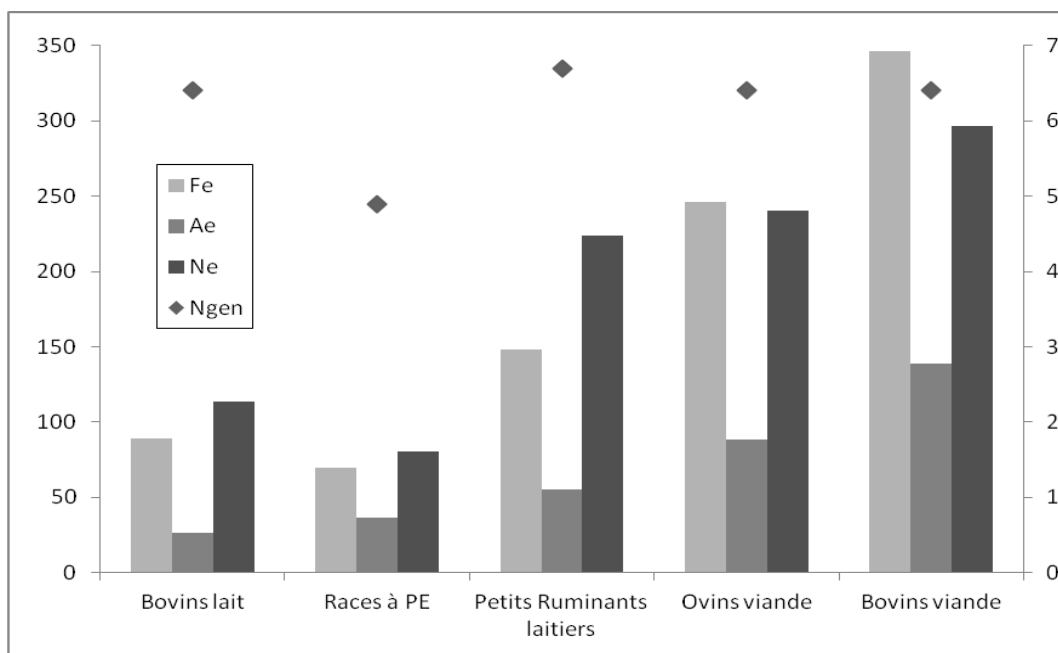


Figure 2 : Indicateurs de variabilité génétique pour les races françaises de ruminants, issus des probabilités d'origine des gènes (Fe, Ae), de la probabilité d'identité des gènes (Ne) [ordonnée de gauche] et de qualité des généalogies (Ngen : nombre d'équivalent génération) [ordonnée de droite].

Nous pouvons illustrer cela dans le cas des filières laitières où l'intensité de sélection est très importante. Les indicateurs de variabilité génétique sont inversement proportionnels au pouvoir de diffusion de l'IA, ils sont donc :

- Les plus faibles pour les bovins, où l'on peut collecter des dizaines de milliers de doses pour un seul taureau. C'est aussi dans cette filière que le taux de pénétration de l'IA est le plus élevé ;
- Meilleurs pour les caprins, où le nombre de doses d'IA est limité à quelques centaines par mâle, mais où l'utilisation de la semence congelée est possible (ce qui signifie que l'on peut constituer des stocks : on augmente ainsi le pouvoir de diffusion d'un mâle dans le temps et dans l'espace, la semence pouvant être distribuée sur l'ensemble du territoire français, voire à l'étranger) ;
- Les plus élevés pour la Lacaune, une race ovine. C'est une espèce où le pouvoir de diffusion est limité à la fois par le nombre de doses que l'on peut collecter par mâle (similaire à celui des caprins) et par une utilisation en semence fraîche. La cryoconservation de la semence congelée est possible mais cette technique n'est pas utilisée au quotidien car elle nécessite, au moment de l'insémination, une opération chirurgicale, la laparoscopie, pour obtenir une fécondité suffisante.

Les derniers facteurs explicatifs de la variation des indicateurs sont l'existence d'une gestion de la diversité génétique dans la race et de flux de gènes étrangers (éventuellement au sens premier du terme, c'est à dire provenant d'animaux sélectionnés en dehors de la France). Par exemple, au sein des races bovines laitières (Danchin-Burge *et al.*, 2011), celles qui montrent les indicateurs de variabilité les plus élevés sont celles qui ont fait varier l'origine des reproducteurs d'insémination au cours du temps.

Pour les races allaitantes, les indicateurs élevés de variabilité des races bovines sont liés à la conjugaison entre effectifs importants et diffusion limitée des mâles du noyau de sélection dû, entre autres, à une faible utilisation de l'IA. Les petits ruminants présentent une plus grande variété de situation entre effectifs et poids du noyau de sélection dans la diffusion des reproducteurs.

Pour les races à petits effectifs, les situations favorables sont dues à une absence de sélection, l'existence éventuelle de programme de gestion de la diversité et un ratio élevé entre nombre de reproducteurs mâles et nombre de reproducteurs femelles (Danchin-Burge *et al.*, 2010). Parmi ces races, les situations à problèmes conjuguent en général effectifs limités, pratiques d'accouplement entre apparentés proches et/ou nombre de fondateurs limités à la relance du programme de conservation. On note qu'en moyenne les indicateurs issus de la probabilité d'origine des gènes sont meilleurs que chez les bovins laitiers : l'absence de sélection dans les races à petits effectifs limite la création de goulets d'étranglement. En revanche, la taille limitée de leurs effectifs implique un accroissement plus rapide de la consanguinité que dans les races bovines laitières, ce qui se traduit par un effectif génétique moyen inférieur pour les races à petits effectifs que pour les bovins laitiers.

2.2. Données moléculaires

Jusqu'à maintenant, le principal frein à l'utilisation des données moléculaires pour caractériser la variabilité génétique d'une population était le nombre limité de marqueurs moléculaires à disposition. Les progrès récents réalisés de la connaissance du génome avec la possibilité de caractériser plus finement des animaux grâce à une densité de marqueurs moléculaires de plus en plus importante (de panels de quelques microsatellites à des puces à 54k SNP voire désormais à 800k SNP) et à des coûts raisonnables, ont permis de créer une base très fournie d'information potentielle sur la variabilité génétique des individus et des populations. De nombreux projets actuels utilisent ces données pour faire de la sélection génomique, trouver de nouveaux QTL ou des signatures de sélection (PhénoFinlait, SheepSNPQTL, GEMBAL, Roquefort-in, etc.), améliorer la connaissance du génome des différentes espèces... Or, ces marqueurs sont également une source particulièrement intéressante pour caractériser la variabilité génétique, mais elle n'est que peu exploitée dans ce sens : la plupart des travaux focalisés sur la diversité des races ont en réalité comme objet d'étude leur histoire génétique

(Groeneveld *et al.*, 2010 ; The Bovine HapMap Consortium *et al.*, 2009 ; Laloë *et al.*, 2010) pour les bovins et (Kijas *et al.*, 2009) pour les ovins.

Les indicateurs calculés à partir des données de génotypages sont principalement liés à la consanguinité, l'apparentement et de façon liée la taille efficace des populations. On appelle consanguinité moléculaire l'indicateur suivant (Baumung *et al.*, 2003) :

$$\frac{1}{n} \sum_{L=1}^n \left(1 - \frac{H_o}{H_e}\right)$$

avec H_o , le taux d'hétérozygotie observée et H_e , le taux d'hétérozygotie attendue.

Cet indicateur est à manier avec précaution car son interprétation est en fait assez différente de celle de la consanguinité calculée à partir des généalogies. Les indicateurs basés sur le calcul d'une « parenté moléculaire » semblent plus pertinents, mais demandent en général un accompagnement pour leur interprétation : il est fréquent que l'on trouve des valeurs négatives qui s'expliquent facilement du point de vue théorique, mais cela reste difficile à vulgariser.

Au final, l'indicateur qui paraît le plus intéressant est celui qui permet de calculer des tailles efficaces de population, et plus particulièrement celui qui est basé sur les déséquilibres de liaison. On appelle déséquilibre de liaison le fait que deux allèles, appartenant à deux locus différents sur un même chromosome, se retrouvent associés de manière plus fréquente que par simple hasard. Ce phénomène laisse supposer une association des allèles due à une proximité physique entre locus. Son estimation peut être utilisé pour calculer la taille efficace d'une population (Flury *et al.*, 2010 ; Qanbari *et al.*, 2010). La vulgarisation de cet indicateur est assez aisée.

Par ailleurs, une approche rendue possible par les marqueurs moléculaires, et qu'il est impossible de réaliser à partir des généalogies, est la recherche des SNP qui contribuent le plus à la variabilité intra génome, et donc les régions génomiques qui concourent le plus à la variabilité intra race. Cela s'avère particulièrement intéressant pour échantillonner des animaux à destination de collections de matériel biologique conservé *ex situ*, comme les collections de la Cryobanque nationale (www.cryobanque.org), afin de conserver le plus de diversité possible dans les collections. Un travail préliminaire réalisé par Engelsma *et al.* (2011) montre que cette approche est à privilégier principalement quand les généalogies sont absentes ou peu fiables ou si l'on souhaite conserver la diversité de certaines parties spécifiques du génome.

L'utilisation des deux sources d'information, généalogies et données de génotypages, sont assez complémentaires. Les marqueurs utilisés dans le cadre de la sélection génomique portent une information assez liée au polymorphisme sélectionné, tandis que les indicateurs issus des généalogies simulent une transmission neutre du polymorphisme.

3. Le projet VARUME

Face aux besoins des filières, une réflexion menée par l'Institut de l'Élevage, avec la participation de l'INRA, l'UNCEIA (Union Nationale des Coopératives Agricoles d'Élevage et d'Insémination Animale), Races de France et l'IFCE (Institut Français du Cheval et de l'Équitation) a abouti à la rédaction du projet VARUME, qui a été un des lauréats d'un financement du CASDAR « innovation et partenariat » (financement du Ministère chargé de l'Agriculture), pour la période 2012-2014. L'objectif de VARUME est de mettre en place un observatoire de la variabilité génétique des différentes populations animales sélectionnées de ruminants et d'équidés, afin de produire de façon régulière des indicateurs permettant d'apprécier la variabilité génétique d'une population animale, sur des bases communes reconnues par l'ensemble des gestionnaires.

3.1. Mise en place d'un observatoire basé sur les données généalogiques

3.1.1. Choix du logiciel de travail

Plusieurs logiciels permettant d'effectuer différentes analyses généalogiques sont disponibles en libre accès. Parmi ceux-ci, citons :

PEDIG (Boichard, 2002) <http://www4.jouy.inra.fr/gabi/les-Ressources/Developpement-d-outils/Pedig>

et ENDOG (Gutiérrez et Goyache, 2005) http://www.ucm.es/info/prodanim/html/JP_Web.htm

Par rapport à PEDIG, ENDOG présente une plus grande convivialité et fournit des sorties plus complètes. En revanche, ENDOG présente deux limites importantes. D'une part, il ne peut pas traiter de fichier comprenant plus que quelques dizaines de milliers de données. Or, pour les races traitées, certaines possèdent des fichiers généalogiques de plusieurs millions d'individus (jusqu'à une vingtaine de millions en race bovine Charolaise et Prim'Holstein), hors de portée donc du logiciel ENDOG. D'autre part, ce logiciel ne fonctionne que si un certain nombre de règles sont strictement respectées, notamment celle qui veut qu'un reproducteur ait au minimum un an de plus que sa progéniture. Or, pour les petits ruminants, il est possible qu'un reproducteur et son ou ses premier(s) descendant(s) naissent la même année.

3.1.2. Les populations analysées

Une première réflexion a permis de définir le périmètre des bilans de variabilité génétique qui seront publiés. Il a été convenu que l'objet d'étude sera la race (en fonction des définitions existantes), avec un cas particulier, celui de la race ovine Lacaune laitière et à viande. En raison de la divergence forte de sélection des différents rameaux, il sera édité un rapport par race (Lacaune lait, Lacaune viande), puis un rapport par entreprise de sélection et par filière.

Les différents indicateurs seront calculés pour une « population analysée ». Cette population a été définie comme, par race, pour les ruminants, (1) les femelles nées pour les 4 dernières années de naissance disponibles ; (2) tous les mâles d'IA qui sont au moins une fois père de la population femelle analysée. Pour les équidés, l'intervalle des 4 dernières années est également retenu, et la population analysée sera constituée à la fois des mâles (entier ou hongre) et des femelles. Ce choix pour les équidés a été nécessaire étant donné la difficulté à définir la population de reproducteur dans ces espèces.

Les résultats des calculs des indicateurs de variabilité génétique seront publiés à condition que l'on dispose d'un minimum d'information généalogique. Ce niveau « suffisant » a été fixé à 2,5 générations connues en moyenne pour la population analysée.

3.1.3. Origine des données généalogiques

Pour les ruminants, toutes les données généalogiques proviennent des Systèmes Nationaux d'Informations Génétiques (SNIG) et sont stockées sur des serveurs du Centre de Traitement de l'Information Génétique (CTIG) de l'INRA. Pour les équidés, les informations sont issues du Système d'Information Relatif aux Equidés (SIRE) géré par l'IFCE. Concrètement, ces choix signifient que les données traitées ne sont pas restreintes aux animaux reproducteurs ni aux animaux enregistrés dans un système de contrôle de production (lait ou viande).

Au final, différents indicateurs seront calculés et publiés pour les populations analysées. Ces indicateurs seront statiques (comme par exemple le coefficient de consanguinité moyen de la population analysée, l'effectif génétique ...) ou évolutifs (par exemple l'évolution de la consanguinité et du niveau d'information généalogique (EqG) pour les populations).

3.2. *Faisabilité d'un observatoire basé sur les données moléculaires*

Dans cette partie du projet, l'objectif est d'utiliser des données de génétique moléculaire existantes pour enrichir nos connaissances sur la variabilité génétique inter et intra-race. L'objectif est, après de premières études sur quelques cas types, de choisir les indicateurs les plus pertinents et d'étudier la faisabilité de la mise en place d'un observatoire de la diversité à partir de données moléculaires en routine. Une comparaison des résultats obtenus pour les races communes aux deux observatoires (moléculaires et généalogiques) est également réalisée. Enfin, il s'agira de définir un protocole permettant de repérer les animaux originaux par leur génome, pour une éventuelle intégration en Cryobanque nationale. Les jeux de données (VALOGENE et ROQUEFORT'IN) sont issues de la sélection génomique pour les races bovines et ovines laitières sélectionnées en France.

Une première étude (Diguët, 2012) a permis d'obtenir une liste de programmes permettant de lire les données de génotypage, de les formater, puis de lancer différents modules de calculs d'indicateurs de variabilité génétique. L'ensemble n'a été testé que pour deux races (Montbéliarde et Lacaune) mais pour deux filières différentes (respectivement bovine laitière et ovine laitière). Le calcul de ces indicateurs pour d'autres races devrait donc être assez aisé par la suite.

Une autre suite attendue du travail est d'étudier la robustesse des indicateurs calculés à partir des données moléculaires, en particulier la taille efficace de la population, en testant différentes méthodes d'échantillonnage pour limiter les variations trouvées quand l'indicateur est calculé par période de naissance.

3.3. *S'assurer de la compréhension des indicateurs*

Il paraît indispensable que la publication d'indicateurs soit accompagnée d'explications sur la façon dont ils sont calculés. En effet, en fonction des informations publiées, il pourra s'avérer nécessaire dans certaines races de changer la façon dont la variabilité génétique est prise en compte dans le programme de sélection ou de conservation. Mais pour que les bonnes décisions soient prises, il faut s'assurer de la bonne compréhension des indicateurs par les gestionnaires.

En particulier, il est indispensable que les gestionnaires aient en tête que les indicateurs dépendent du niveau d'information généalogique. Par ailleurs, l'expérience montre que la présentation de la liste des ancêtres majeurs est toujours sujette à discussion, car cela correspond rarement à la vision que les gestionnaires ont de la race. Enfin, à l'usage, l'indicateur qui nécessite le plus d'explication est, paradoxalement, un de ceux qui est le plus familier aux éleveurs et aux responsables de leurs organisations, à savoir la consanguinité. Il est toujours nécessaire de définir la consanguinité en lien avec la parenté, tout en insistant sur les différences entre ces deux concepts : la consanguinité concerne un animal seul, la parenté concerne des paires d'individus ; la parenté se transmet à la descendance, pas la consanguinité ; etc. Il apparaît également utile d'insister sur le fait que c'est plus le rythme d'accroissement de la consanguinité (par exemple traduit par un effectif génétique) que son niveau absolu qui a du sens en matière de suivi de variabilité génétique.

Ce besoin de vulgarisation a été bien perçu et c'est pour cela que des formations spécifiques ainsi que la rédaction d'un guide de lecture des indicateurs seront mises en place dans le cadre de VARUME.

Conclusion

Il existe des besoins évidents et des demandes fortes dans les différentes filières pour disposer d'indicateurs régulièrement mis à jour. Ces besoins se sont accrus avec l'avènement de la sélection génomique, qui va raccourcir les intervalles entre générations. Il devient crucial d'évaluer rapidement l'impact des nouveaux programmes sur la variabilité génétique des races et s'assurer ainsi de leur durabilité. La production en routine d'indicateurs d'évolution de la variabilité génétique permettra aux

organismes gestionnaires des programmes de sélection et de conservation (en particulier les Organismes et Entreprises de Sélection) d'être alertés sur la « santé » génétique de leur population et d'adapter leurs pratiques de gestion de la variabilité génétique.

La construction de l'observatoire de la variabilité génétique permet à la France de consolider ses engagements à ce titre, dans la continuité des actions déjà existantes à ce niveau comme le suivi de programmes de conservation par l'Institut de l'Élevage (ruminants) et l'IFCE (équidés) en collaboration avec l'INRA, Races de France et les associations de races concernées (*in situ*), et la création de la Cryobanque nationale (collection patrimoniale de matériel biologique – *ex situ*) en 1999.

Références bibliographiques

Avon L., Colleau J.J., 2006. Conservation in situ de 11 races bovines françaises à très faibles effectifs: bilan génétiques et perspectives. *Rencontre Recherche Ruminants* 13, 247-250.

Baumung R., Sölkner J., 2003. Pedigree and marker information requirements to monitor genetic variability. *Genetics Selection Evolution* 35, 369-383.

Boichard D., 2002. PEDIG : a Fortran package for pedigree analysis suited for large populations. *Proceedings of the 7th WCGALP*, Montpellier, France. Communication No. 28-13.

Boichard D., Maignel L., Verrier E., 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution* 29, 5–23.

Bokor Á., Jónás D., Ducro B., Nagy I., Bokor J., Szabari M. 2013. Pedigree analysis of the Hungarian Thoroughbred population. *Livestock Science*. 151, 1-10.

Cervantes I., Goyache F., Molina A., Valera M., Gutiérrez J.P., 2011. Estimation of effective population size from the rate of coancestry in pedigreed populations. *Journal of Animal Breeding and Genetics* 128, 56-63

Danchin-Burge C., Avon L., 2000. Analyse de la variabilité génétique de races bovines à très petits effectifs, après vingt ans d'actions de conservation. *Rencontre Recherche Ruminants* 7, 145-148.

Danchin-Burge C., Palhière I., François D., Bibé B., Leroy G., Verrier E., 2010. Pedigree analysis of seven small French sheep populations and implications for the management of rare breeds. *Journal of Animal Sciences* 88, 505-516.

Danchin-Burge C., Leroy G., Brochard M., Moureaux S., Verrier E., 2011. Evolution of the genetic variability of eight French dairy cattle breeds assessed by pedigree analysis. *Journal of Animal Breeding and Genetics* 129, 206–217.

Danchin-Burge C., Allain D., Clément V., Piacère A., Martin P., Palhière I., 2012. Genetic variability and French breeding programs of three goat breeds under selection. *Small Ruminant Research* 108, 36-44.

Danvy S., Heydeman P., 2005. Le point sur les ressources génétiques équine en France. *Journées de la Recherche Equine* 31, 135-146.

Diguet J., 2012. Faisabilité d'un observatoire de la variabilité génétique des ruminants à partir de données moléculaires. Mémoire de fin d'études, Bordeaux Sciences Agro, Bordeaux, France, 55 p.

Engelsma KA, Veerkamp RF, Calus MP, Windig JJ., 2011. Consequences for diversity when prioritizing animals for conservation with pedigree or genomic information. *Journal of Animal Breeding and Genetics*. 128(6):473-81.

Falconer D.S., 1981. Introduction à la génétique quantitative. Masson, Paris.

Flury C., Tapio M., Sonstegard T., Drögemüller C., Leeb T., Simianer H., Hanotte O., Rieder S., 2010. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium, *Journal of Animal Breeding and Genetics* 127, 339-347.

Groeneveld L.F., Lenstra J.A., Eding H., Toro M.A., Scherf B., Pilling D., Negrini R., Finlay E.K., Jianlin H., Groeneveld E., Weigend S. and The GLOBALDIV Consortium, 2010. Genetic diversity in farm animals – a review. *Anim. Genet.* 41 (Suppl. 1), 6–31.

- Gutiérrez J.P., Goyache F., 2005. ENDOG : a computer program for monitoring genetic variability of populations using pedigree information. *Journal of Animal Breeding and Genetics* 122, 357-360.
- Institut de l'Élevage, 2009. Estimation de la variabilité génétique de 19 races bovines à partir de leurs généalogies. *Compte rendu d'étude final n° 0009 72 125*, 74 p.
- Institut de l'Élevage, 2010. Bilan de variabilité génétique de 35 races ovines. *Compte rendu d'étude final n°0010 72 045*, 118 p.
- Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F., McGrath A., Wilson P., Ingersoll R.G., McCulloch R., McWilliam S., Tang D., McEwan J., Cockett N., Oddy V.H., Nicholas F.W., Raadsma H., 2009. A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLoS ONE* 4(3): e4668. doi:10.1371/journal.pone.0004668.
- Koenig S., Simianer H., 2006. Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. *Livestock Sciences*, 103, 40–53.
- Laloë D., Moazami-Goudarzi K., Lenstra J.A., Marsan P.A., Azor P., *et al.*, 2010. Spatial Trends of Genetic Variation of Domestic Ruminants in Europe. *Diversity* 2, 932–945.
- Leroy G., Mary-Huard T., Verrier E., Danvy S., Charvolin E., Danchin-Burge C., 2012. Estimating the effective population size using pedigree data. What method should be used in practice? Examples in dogs, sheep, cattle and horses. *Genetics Selection Evolution* 45, 1. doi:10.1186/1297-9686-45-1.
- Mc Parland, S., Kearney J. F., Rath M., Berry D. P., 2007. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. *Journal of Animal Sciences* 85, 322–331.
- Piacère A., Palhière I., Rochambeau H. de, Allain D., 2004. Analysis of the genetic variability of the French Alpine and Saanen breeds using genealogical data. *Conference on Goats 4-9 July 2004, South Africa*.
- Poncet P. A., Pfister W., Muntwyler J., Glowatzki-Mullis M. L., Gaillard C., 2006. Analysis of pedigree and conformation data to explain genetic variability of the horse breed Franches-Montagnes. *Journal of Animal Breeding and Genetics*. 123, 114–121.
- Qanbari S., Hansen M., Weigend S., Preisinger R., Simianer, H., 2010. Linkage disequilibrium reveals different demographic history in egg laying chickens, *BMC Genetics* 11:103 doi:10.1186/1471-2156-11-103.
- Rizzi E., Tullo A., Cito M., Caroli A., Pieragostini E., 2011. Monitoring of genetic diversity in the endangered Martina Franca donkey population. *Journal of Animal Sciences* 89, 1304–1311.
- Selvaggi M., Dario C., Peretti V., Ciotola F., Carnicella D., Dario M., 2010. Inbreeding depression in Lecce sheep. *Small Ruminant Research* 89, 42-46
- The Bovine HapMap Consortium, Gibbs R.A., Taylor J.F., Van Tassell C.P., Barendse W., Eversole K.A., Gill C.A., Green R.D., Hamernik D.L., Kappes S.M. *et al.*, 2009. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* 324, 528-532.
- Valera M., Molina A., Gutiérrez J.P., Gómez J., Goyache F., 2005. Pedigree analysis in the Andalusian horse : population structure, genetic variability and influence of the Carthusian strain. *Livestock Production Science*. 95, 57–66.
- Verrier E., Colleau J.J., Foulley J.L., 1993. Long term effects of selection based on the animal model BLUP in a finite population. *Theor. Appl. Genet.* 87, 446-454.