



Création d'un observatoire de la VARIabilité génétique des RUMinants et des Equidés (VARUME)

C. Danchin-Burge, S. Danvy, Denis Laloë, Etienne Verrier

► To cite this version:

C. Danchin-Burge, S. Danvy, Denis Laloë, Etienne Verrier. Création d'un observatoire de la VARIabilité génétique des RUMinants et des Equidés (VARUME). *Innovations Agronomiques*, 2017, 55, pp.235-245. 10.15454/1.5137783633347627E12 . hal-01604871

HAL Id: hal-01604871

<https://hal.science/hal-01604871>

Submitted on 25 May 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.



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

Création d'un observatoire de la VARIabilité génétique des RUMinants et des Equidés (VARUME)

Danchin-Burge C.¹, Danvy S.², Laloë D.³, Verrier E.³

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

² IFCE, Institut français du cheval et de l'équitation, Avenue de l'école nationale d'équitation, BP207 – Terrefort 49411 Saumur

³ INRA/AgroPariTech, UMR1313 GABI, 78350, Jouy-en-Josas.

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 l'utilisation de données de génotypes pour réaliser 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 y est corrélé, 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 génotypes. 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 « santé génétique » des races, sur des bases communes reconnues par l'ensemble des gestionnaires.

Mots-clés : indicateurs, généalogies, données moléculaires, races menacées, races en sélection.

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

Selection programs in livestock species are becoming more and more efficient. New tools, such as molecular data used for genomic selection is accelerating how livestock breeds are evolving. Yet it is necessary to preserve the breeds' genetic variability firstly because genetic progress is correlated to it, and any sudden increase of the inbreeding level can impact these populations in a negative way. Therefore it is necessary to regularly assess the genetic variability of both rare breeds and selected breeds in order to maintain their sustainability. Two types of data make it possible to calculate genetic variability indicators: pedigrees and genotyped data. For instance, thanks to these indicators, we can evaluate the impact of a major ancestor on the current population. The main goal of the VARUME project was to set up a genetic variability observatory for ruminant and equid breeds, in order to provide, on a regular basis, indicators that can assess the "genetic health" of the breeds, and that are agreed on by all the breed managers.

Keywords: indicators, pedigrees, molecular data, rare breeds, selected breeds.

Introduction

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 peu nombreuse de reproducteurs. 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 qui 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 nécessitent 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.

Il existe des besoins évidents et des demandes fortes de la part des gestionnaires de races menacées et des Organismes de Sélection dans les différentes filières pour disposer d'indicateurs régulièrement mis à jour. Ces besoins se sont accrus d'autant qu'avec l'avènement de la sélection génomique, qui va raccourcir les intervalles entre générations, il est 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 permettrait 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.

Face aux besoins des filières, une réflexion menée par l'Institut de l'Élevage, avec la participation de l'INRA, ALLICE, 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. Deux types de données sont utilisés pour créer cet observatoire, les données des généalogies et les données moléculaires.

1. Démarche utilisée

Si la variabilité génétique se réfère à une diversité d'allèles portés par les animaux d'une population, l'information génomique ne constitue pas l'unique source de données utilisée pour caractériser cette variabilité. Tant qu'il était difficile d'observer un grand nombre de locus, d'autres sources de données étaient mobilisées, à savoir les phénotypes et les généalogies (Verrier et al., 2005). Cependant, avec la révolution génomique et l'automatisation des techniques de génotypage, il est désormais possible d'obtenir une information génomique en quantité et à coût réduit.

Pour le projet VARUME, nous nous sommes focalisés sur l'utilisation des données généalogiques et moléculaires, de type SNP (« Single Nucleotide Polymorphism ») pour évaluer la variabilité génétique des races de ruminants et d'équidés. Les deux paragraphes suivants précisent la démarche d'abord dans le cadre des généalogies puis des données moléculaires.

1.1 Indicateurs de variabilité génétique basés sur les généalogies

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 (Koenig et 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).

1.1.1 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 limitées aux seuls animaux reproducteurs. Une restriction existe cependant pour les ovins et les caprins, où seuls les animaux au contrôle officiel sont renseignés.

1.1.2 Populations analysées

L'objet d'étude est la race en tant que telle et dans son ensemble (en suivant la réglementation existante). Les différents indicateurs sont calculés pour une « population analysée ». Pour chaque race de ruminants, une année donnée, cette population est composée, d'une part, de toutes les femelles nées lors des 4 dernières années et, d'autre part, de tous les mâles d'insémination (IA) qui sont au moins une fois père dans la population des femelles analysées. Pour les équidés, l'intervalle des 4 dernières années est également retenu. La population analysée est constituée à la fois des mâles (entier ou hongre) et des femelles. Ce choix pour les équidés a été motivé par la difficulté, dans cette famille d'espèces, à définir la population des reproducteurs.

1.1.3 Logiciel de travail

Plusieurs logiciels permettant d'effectuer différentes analyses généalogiques sont disponibles en libre accès. Pour développer les indicateurs de VARUME, c'est le logiciel PEDIG (Boichard, 2002), libre d'usage, qui a été utilisé. En dehors des critères démographiques, tous les autres indicateurs présentés dans les paragraphes suivants sont issus de méthodes de calcul utilisées par ce logiciel et présentées pour la plupart par Boichard et al. (1997).

1.1.4 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), il est alors non apparenté au reste de la population. La qualité des généalogies est appréciée par un indicateur appelé « nombre équivalent de générations connues » (Ngen), dont le calcul nécessite de répertorier les ancêtres connus sur chaque génération. Cet indicateur s'interprète comme suit : la qualité des généalogies de la population analysée est équivalente à celle d'une population où, sur un nombre de générations égal à Ngen, tous les ancêtres seraient connus. Par exemple, une race avec un Ngen de 5 présente des généalogies d'une qualité équivalente à celle d'une population dont tous les ancêtres seraient connus sur 5 générations. Les résultats des calculs des indicateurs de variabilité génétique sont publiés à condition que l'on dispose d'un minimum d'information généalogique. Ce niveau « suffisant » a été fixé à une moyenne de 2,5 générations connues d'ancêtres pour la population analysée.

1.1.5 Indicateurs démographiques

Des indicateurs démographiques simples (comme le nombre de pères et de grands-pères maternels dont est issue la population analysée) sont des notions très parlantes pour les gestionnaires. De plus, ces informations permettent d'apprécier simplement la variabilité génétique pour des races dont les généalogies sont mal connues, et donc pour qui les indicateurs déduits de ces généalogies peuvent être considérés comme modérément fiables. Ces indicateurs sont calculés grâce à des statistiques simples développées spécifiquement dans le cadre du projet VARUME.

1.1.6 Indicateurs issus de la probabilité d'origine des gènes : les ancêtres majeurs

On utilise l'approche des probabilités d'origine des gènes afin de détecter 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 – appelée contribution –, on considère :

- que les fondateurs (f, ancêtres eux-mêmes sans parents connus) sont la source de tous les gènes actuels (pas de prise en compte de mutations éventuelles).
- qu'un gène d'un individu 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.

Pour la population analysée, on calcule donc les probabilités d'origine des animaux étudiés, c'est-à-dire la contribution de tous les ancêtres : parents, grands-parents etc. On obtient une liste d'ancêtres, avec la contribution de chacun à la population étudiée. Ces contributions sont appelées contributions brutes. L'inconvénient de ces contributions est qu'elles représentent mal l'importance des différents ancêtres, car en général ils sont apparentés entre eux : par exemple la contribution d'un mâle n'est pas indépendante de celle de son père ou de son fils. Pour s'affranchir de ce problème, on calcule dans un second temps, une contribution marginale, c'est-à-dire sa contribution à la population, indépendamment des autres animaux qui lui sont apparentés. A partir des contributions, on calcule un nombre efficace d'ancêtres (Ae) correspondant 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 sont moins sensibles que d'autres à la 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.

L'approche des probabilités 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 où une origine est attribuée à chaque fondateur (c'est à dire un animal sans parents connus) et l'évolution de la transmission des gènes est suivie d'une génération à l'autre par cette approche probabiliste.

1.1.7 Indicateurs issus de la probabilité d'identité des allèles : parenté et consanguinité

La question sous-jacente au calcul de ces indicateurs est de savoir si les allèles portés par un individu sont identiques par descendance, ce qui peut se produire s'il est consanguin. Un animal est consanguin si ses parents sont apparentés entre eux. Deux animaux sont apparentés s'ils possèdent au moins un ancêtre commun.

Considérons un animal consanguin et un gène quelconque chez cet animal. Les deux parents de celui-ci ayant un ancêtre commun, ils peuvent, pour le gène considéré, avoir reçu chacun la copie d'un même allèle présent chez cet ancêtre commun. Il est donc possible qu'ils transmettent chacun cette copie de l'allèle reçue de l'ancêtre commun : ainsi, leur descendant (consanguin) peut avoir, en un gène quelconque, un génotype constitué de deux copies du même allèle ancestral. Deux allèles issus de la

duplication du même allèle ancestral sont dits « identiques par descendance ». Le coefficient de consanguinité d'un animal est égal à la probabilité que, pour un gène quelconque, les allèles portés par cet animal soient identiques par descendance. On montre que ce coefficient de consanguinité du descendant est égal au coefficient de parenté entre ses deux parents. On peut ensuite généraliser les calculs individuels de consanguinité à l'ensemble d'une race.

Pour illustrer de manière plus parlante ce que représente l'accroissement de la consanguinité sur une période donnée, et pouvoir comparer plus aisément les races entre elles, il existe un indicateur appelé « taille efficace » (notée « Ne » et également appelé « effectif génétique »). Une population de taille efficace « Ne » a le même accroissement de consanguinité ou de parenté qu'une population qui a des effectifs de la valeur « Ne » mais qui se reproduit, elle, en conditions « idéales », c'est à dire sans sélection, et avec des accouplements aléatoires. Il s'agit d'un nombre unique, équivalent à un effectif de reproducteurs (mâles et femelles confondus), prenant en compte les distorsions de taille de descendance qui peuvent exister d'un reproducteur à l'autre. Par exemple, l'accroissement de consanguinité d'une population de taille efficace de 100 correspond à celui d'une population dont l'effectif femelle serait de taille infinie mais dont l'effectif mâle serait de 25 individus seulement. Pour les gestionnaires de race, l'intérêt principal de cette mesure est de suivre son évolution dans le temps.

Il existe plusieurs méthodes de calcul de la taille efficace d'une population à partir des généalogies. Les résultats peuvent être sensiblement différents selon la méthode employée, en particulier pour les races dont l'accroissement de consanguinité est faible ou très variable d'une année sur l'autre (Leroy et al., 2013). Au vu des connaissances actuelles, la méthode la plus pertinente, mise en œuvre dans le cadre de VARUME, est une méthode fondée sur l'accroissement de la parenté et prenant en compte la connaissance des généalogies animal par animal.

1.2 Indicateurs issus des génotypages

Jusqu'au début des années 2010, 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 ont émergé dont l'objectif était d'utiliser 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 ; en bovins : The Bovine HapMap Consortium et al., 2009 et Laloë et al., 2010 ; en ovins : Kijas et al., 2009).

1.2.1 Origine des données moléculaires

Les méthodes permettant l'analyse de la structure démographique d'une population à partir de données de génotypage existent depuis les années 1970. Elles ont connu un regain d'intérêt avec l'apparition des génotypages à haut débit, en particulier les puces SNP (« Single Nucleotide Polymorphism »). Ces puces regroupent des dizaines de milliers, voire des centaines de milliers de marqueurs SNP. Ces derniers sont des simples nucléotides, localisés précisément sur les chromosomes. Les SNP ont en règle générale deux allèles seulement.

Les données moléculaires utilisées dans les analyses proviennent de différents consortiums (Roquefort'IN et GENOMIA pour les données ovines ; AMASGEN puis VALOGENE pour les données bovines).

1.2.2 Choix des indicateurs calculés à partir de génotypages

Différents indicateurs reflétant la diversité génétique peuvent être obtenus à partir de génotypages, telles que la richesse allélique - peu informative pour des marqueurs bialléliques comme les SNP - ou le nombre moyen des marqueurs polymorphes (hétérozygotie). Il est difficile pour ce dernier indicateur de s'abstraire de la taille de l'échantillon typé. C'est pourquoi nous proposons l'utilisation de la taille efficace de la population estimée à partir du déséquilibre de liaison (DL), pour laquelle existe une importante littérature scientifique.

La notion fondamentale sur laquelle sont fondées les études de démographie et de variabilité génétique est le « déséquilibre de liaison » (DL). Formellement, le DL est défini comme l'existence d'une association préférentielle entre allèles à deux marqueurs différents. En règle générale, on mesure ce déséquilibre par le carré de la corrélation (r^2) entre les deux marqueurs. Ce coefficient varie entre 0 à l'équilibre de fréquences et 1 quand le déséquilibre est total, c'est-à-dire quand un allèle du premier marqueur est systématiquement associé à un allèle du second (c'est-à-dire quand il n'existe que deux combinaisons alléliques, par exemple AB et ab seulement). Au final, la taille efficace des populations étudiées a été calculée en utilisant la formule suivante (Waples, 2006) :

$$N_e = \frac{1}{3(r^2 - \frac{1}{n})}$$

où N_e est la taille efficace, r^2 le DL, et n la taille de l'échantillon.

2. Suivre l'évolution de la variabilité génétique des races de ruminants et d'équidés grâce à l'observatoire VARUME

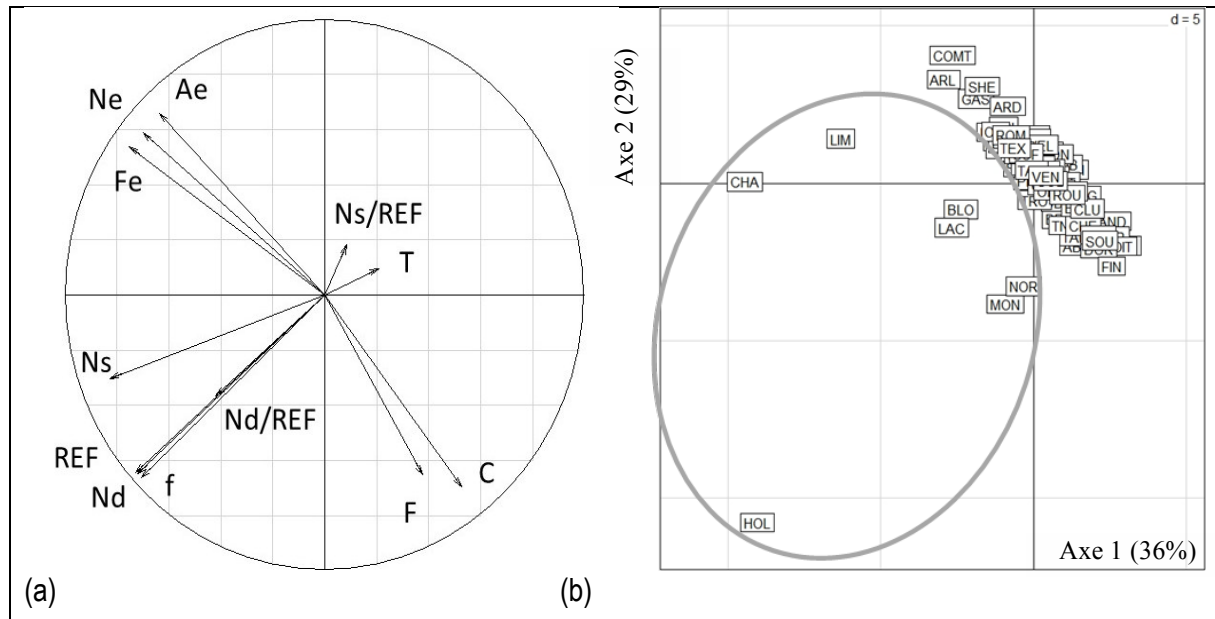
2.1 Corrélations entre les indicateurs généalogiques

Nous avons réalisé une Analyse en Composantes Principales (ACP) pour toutes les races avec certains des indicateurs de variabilité génétique calculés en 2014 (Figure 1a), qui ont été choisis comme les plus représentatifs parmi les critères démographiques (REF: taille de population, Nd: nombre de mères, Ns: nombre de pères, f: nombre de fondateurs), les critères issus de la probabilité d'origine des gènes (Fe: nombre de fondateurs efficaces, Ae: nombre d'ancêtres efficaces) et les critères issus de la probabilité d'identité (F: coefficient de consanguinité, C: coefficient de parenté, Ne: taille efficace de population).

Les deux premiers axes de l'ACP expliquent environ 65% de la variation totale (Figure 1a). Ils permettent la différenciation d'une part des paramètres démographiques (comme REF, Ns, Ns et f, qui ont des corrélations entre eux de plus de 73%), et, d'autre part, les indicateurs de variabilité génétique (C, F, Ae, fe, Ne). Parmi les paramètres démographiques, un résultat intéressant est la corrélation positive significative entre REF et Nd/REF (0.41). Il semble que le ratio de Nd/REF croît plus vite que REF quand la taille de la population augmente. Une interprétation serait que dans les grandes populations, les femelles ont une taille de descendance inférieure à celle que l'on a dans les plus petites populations, en raison d'une pression de sélection plus forte qui conduit à un renouvellement des générations plus rapide.

Fe et f ne sont pas significativement corrélés (0,09) tandis que la corrélation entre f et REF est proche de 1. Notre interprétation est que la variabilité génétique d'une race ne dépend pas de ses effectifs mais de la sévérité des goulets d'étranglement qu'elle a subis. Ce résultat est bien illustré dans la Figure 1a. Une des découvertes intéressantes de notre étude est qu'il y a une corrélation limitée entre les paramètres démographiques (comme REF, Ns et Nd) et les indicateurs de variabilité génétique (Ne inclus). Les grandes populations peuvent être autant dans une situation difficile que les plus petites du point de vue de la variabilité génétique, et une petite population peut avoir une variabilité génétique supérieure à celle d'une très grande si elle est bien gérée.

A l'inverse, la corrélation est élevée entre tous les indicateurs de variabilité génétique : plus Ae et Fe sont petits, plus la consanguinité est élevée. Expliqué autrement, la consanguinité et la parenté ont tendance à être élevées dans une population avec un faible nombre d'ancêtres majeurs.



(a) REF: taille de population, Nd: nombre de mères, Ns: nombre de pères, f: nombre de fondateurs, Fe: nombre de fondateurs efficaces, Ae: nombre d'ancêtres efficaces, F: coefficient de consanguinité, C: coefficient de parenté, Ne: taille efficace de population

Figure 1: Analyse en Composantes Principales de la variabilité génétique des populations sur le premier plan: (a) cercle des corrélations, (b) positions des races

Les résultats de l'ACP (Figure 1b) permettent également de distinguer clairement les races à grands effectifs, quelle que soit l'espèce (i.e. Blonde d'Aquitaine: BLO, Charolais: CHA, Holstein: HOL, Limousine: LIM, Montbéliarde: MON, Normande: NOR, des races bovines, et la Lacaille: LAC, une race ovine). En revanche il n'est pas possible de trouver d'autres types de différences en fonction de l'espèce ou du type de filière.

2.2 Résultats de l'observatoire en 2014

Le nombre d'ancêtres efficaces (Ae), représentatif des goulets d'étranglement subis par les races, la taille efficace (Ne), témoin de l'accroissement récent de la consanguinité, ainsi qu'un indicateur de qualité, le nombre d'équivalent génération (Ngen), ont été calculés pour 89 races de ruminants et d'équins (36 races ovines allaitantes, 20 races de chevaux, 14 races de petits ruminants laitiers, 10 races de bovins laitiers, 9 races de bovins viande), à partir de la population de chaque race née entre 2011 et 2014. Les résultats sont présentés dans la Figure 2.

Les indicateurs calculés sont de bonne qualité car en moyenne le Ngen dépasse la valeur de 5 (hors chevaux), qui est considérée comme étant la valeur minimale pour estimer correctement les généalogies (Baumung et al., 2002).

Les indicateurs de variabilité présentent un continuum avec, du côté le plus avantage, les races bovines allaitantes, et à l'autre extrémité, les races bovines laitières, alors que la taille moyenne des effectifs pour ces deux filières est comparable. On retrouve ici ce qui a été visible à la Figure 1, à savoir que la taille des effectifs ne permet pas de prédire la variabilité génétique d'une population. Les valeurs des indicateurs de variabilité génétique par filière sont le reflet de plusieurs facteurs : taux de pénétration de

l'IA pour renouveler les reproducteurs, effectif des reproducteurs femelles, existence ou non de systèmes de gestion de la diversité et enfin introgression régulière ou non de gènes étrangers.

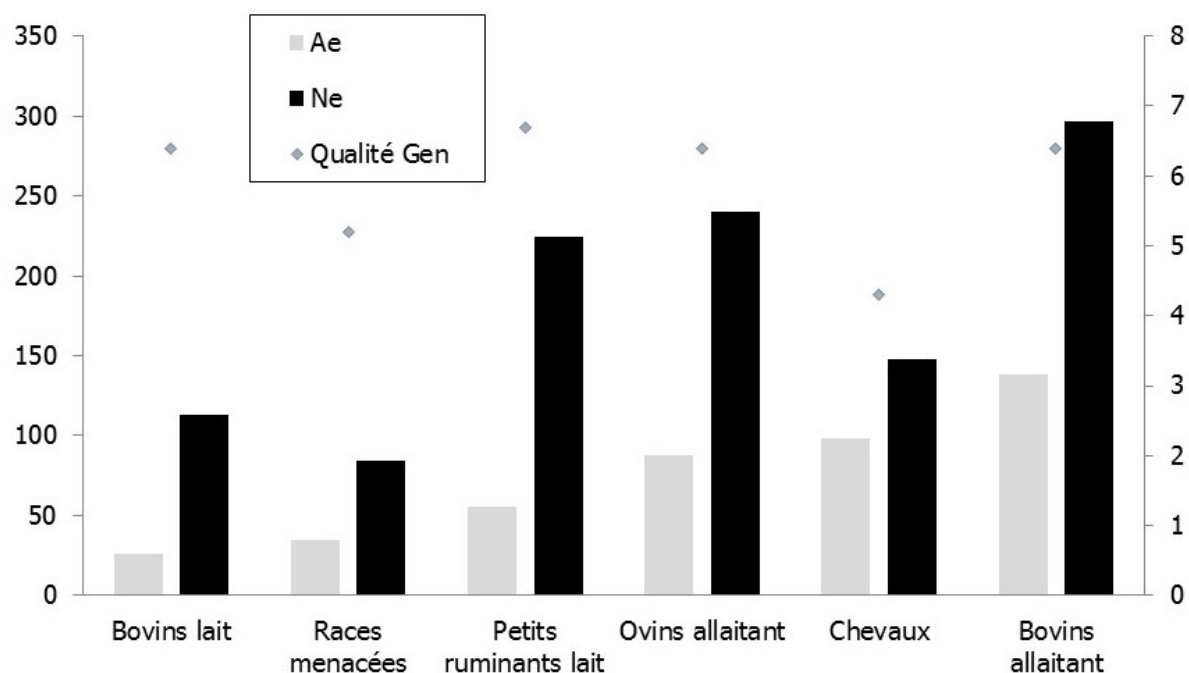


Figure 2 : Indicateurs de variabilité génétique pour les races de ruminants et d'équidés sélectionnés en France, issus des probabilités d'origine des gènes (Ae), de la probabilité d'identité des gènes (Ne) [ordonnée de gauche] et de qualité des généalogies (Ngen : nombre d'équivalent génération) [ordonnée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

Pour illustrer cela, l'intensité de sélection est la plus élevée pour la filière bovine laitière, d'où, en moyenne, un Ae (reflet des goulets d'étranglement) très faible. En revanche pour un critère comme la consanguinité, la taille de la population a plus d'importance, et donc ce sont les races menacées qui ont, en moyenne, les indicateurs de Ne les plus faibles. Pour les races allaitantes, les indicateurs élevés de variabilité des races bovines sont liés à la conjugaison d'effectifs importants et d'une diffusion limitée des mâles du noyau de sélection due, 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.

3. Faisabilité de l'utilisation des données moléculaires pour créer des indicateurs de variabilité génétique

Une première approche a consisté à calculer la taille efficace d'une race bovine laitière, la Normande, à partir de l'ensemble des données disponibles. Pour calculer cet indicateur, après plusieurs essais, il est apparu nécessaire de grouper les données de génotypages par sexe et par période de naissance. En effet, on dispose d'informations pour la quasi-totalité des mâles disponibles pour la reproduction depuis la fin des années 2000, tandis que la population femelle ne représente qu'une faible part de la population non échantillonnée, et on peut s'attendre à un biais d'échantillonnage. Par ailleurs nous disposons de peu d'informations de génotypage avant l'année 2008 (Tableau 1).

Pour la race Normande (Tableau 1), on observe la même tendance entre les deux types de Ne, à savoir une amélioration de cet indicateur (de 66 pour la première période à 73 pour la dernière). Par ailleurs on constate que les valeurs des tailles efficaces des données généalogiques sont à peu près la moitié de

celles des N_e issues des SNP. Ceci contraste avec les résultats des races ovines laitières (Tableau 2). Une explication possible est le fait que le N_e généalogique n'est basé que sur des données relativement récentes (environ 8,5 générations remontées pour les animaux échantillonnés) et il reflète une tendance récente de la variabilité génétique de la Normandie, à savoir une réduction drastique de sa diversité. Les marqueurs moléculaires indiquent eux l'état réel de la population, et notre hypothèse est que le pool fondateur de la Normandie était très variable, ce qui est également signalé par les éléments historiques de la création de la race.

	Données SNP				Pedigree
	Taille échantillon		N_e		N_e
	Male	Femelle	Male	Femelle	Femelle
2005/07	466	3 619	103	132	66
2008/10	2 161	3 996	134	139	68
2011/13	2 251	5 078	151	153	73

Tableau 1 : Estimation de la taille efficace (N_e) basée sur les SNP et les données généalogiques pour la race Normandie, en fonction du sexe et de la période de naissance.

Une seconde série de travaux (François et al., 2014) ont été réalisés à partir des données de SNP provenant de races ovines laitières (Tableau 2). Ils ont montré qu'il était nécessaire de prendre en compte la structuration très particulière de la race Lacaune (coexistence de deux schémas de sélection n'échangeant pas d'animaux depuis plus de 30 ans) pour affiner les résultats. La prise en compte de cette structuration a été faite en utilisant des corrélations partielles pour calculer la taille efficace (N_e LD struct dans le Tableau 2). Avec cette méthode, le N_e obtenu pour la race Lacaune est très proche de celui obtenu avec le N_e généalogique. Pour les deux plus petites races ovines laitières (Basco Béarnaise et Manech Tête Noire), les valeurs obtenues avec le N_e LD « classique » et le N_e généalogique sont déjà très proches.

N_e	N_e LD	N_e LD struct	N_e Pedigree
Lacaune	195	303	312
Manech Tête Rousse	118	145	148
Basco Béarnaise	98	/	91
Manech Tête Noire	92	/	82

Tableau 2 : Taille efficace (N_e) pour 4 races ovines laitières sur la période 2000-2011 (4 générations) avec 3 méthodes différentes

De façon plus générale, l'étude de la faisabilité de l'utilisation de données de type SNP pour calculer des indicateurs de variabilité génétique a été montrée. Pour obtenir une estimation correcte de la taille efficace d'une population à partir de données de type SNP, nous avons trouvé qu'il faut au minimum 60 individus génotypés par période considérée (François et al., 2014), ce qui est corroboré par d'autres études (England et al., 2006 ; Waples, 2006). La population génotypée doit être représentative de la population totale, ce qui limite ces études à l'heure actuelle aux races en sélection génomique. Par ailleurs, pour chaque génération, l'échantillon doit être constitué du même type d'animaux, c'est à dire que si les males seuls ont été génotypés à la première génération, les estimations doivent être basées uniquement sur ce sexe aux générations suivantes.

4. Conclusion

Le projet VARUME permet, depuis 2015, à chaque gestionnaire de race de ruminants et d'équidés (à condition qu'on dispose de suffisamment d'informations généalogiques) d'avoir périodiquement un tableau de bord d'indicateurs de variabilité génétique. Les résultats de l'observatoire sont mis en ligne annuellement sur le site de l'Institut de l'Élevage (<http://www.idele.fr>) pour les races de ruminants. Pour les équidés, les résultats sont diffusés à façon, en fonction des demandes des organismes de sélection.

C'est également un moyen pour la France de consolider ses engagements pris en ratifiant la Convention sur la Diversité Biologique quant à la protection de ses ressources génétiques, et qui suit d'autres actions déjà entreprises comme la création de la Cryobanque nationale (de conservation de matériel biologique d'animaux d'élevage), en 1999, ou bien le suivi des programmes de conservation des races de ruminants par l'Institut de l'Élevage pour les ruminants et par l'IFCE pour les équidés.

Remerciements

Nos remerciements aux consortiums Roquefort'IN et GENOMIA pour la fourniture des génotypes en races ovines laitières et à VALOGENE pour les génotypes bovins lait. Financement du projet par le CASDAR (Ministère chargé de l'Agriculture).

Références bibliographiques

- Baumung R., Sölkner H., 2002. Analysis of pedigrees of Tux-Zillertal, Carinthian Blond and Original Pinzgau cattle population in Austria. *J. Anim. Breed. Genet.* 119, 175–181.
- 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.
- 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.
- England P.R., Cornuet J.-M., Berthier P., Tallmon D.A., Luikart G., 2006. Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Cons. Genet.* 7(2), 303-308.
- 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.
- 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
- 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.
- 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.

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.

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

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.

Waples R., 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Cons. Genet.* 7, 167-184.

Cet article est publié sous la licence Creative Commons (CC BY-NC-ND 3.0)



<https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « *Innovations Agronomiques* », la date de sa publication, et son URL)