

III. L'UTILISATION EN ELEVAGE DES GENES IDENTIFIES ET DES CARACTERES A DETERMINISME SIMPLE

A. Définition et aperçu historique

Les caractères à déterminisme génétique simple sont des caractères pour lesquels peu de locus interviennent, souvent un seul. Ces caractères sont en général peu ou pas influencés par le milieu et on peut faire simplement la liaison entre un allèle donné et la « forme » qu'il détermine. Les gènes en cause induisent des variations phénotypiques importantes entre individus (d'où l'appellation « polymorphisme visible »). En général, ces variations sont discontinues et les individus se répartissent dans des classes.

Historiquement, les premiers efforts de sélection animale et les premières applications de la génétique ont porté sur des caractères à déterminisme simple, notamment les caractères de coloration, du fait de la relative simplicité de leur observation et de leur étude. Les distinctions de couleur du pelage (on parle de robe) ou du plumage ont constitué une forte composante de la définition des standards lorsque, au XIX^{ème} siècle, les éleveurs ont souhaité différencier leurs animaux à partir de la notion de race, les animaux d'une même race ayant des caractéristiques héréditaires communes (voir GQ, chapitre V). La faible liaison qui existe entre beaucoup de caractères de standard et les aptitudes zootechniques a fortement limité l'intérêt de la notion même de standard. Il se trouve cependant que, dans différentes espèces, certains de ces caractères à déterminisme simple revêtent une importance économique (voir plus loin).

Par ailleurs, les outils d'investigation du polymorphisme des populations animales se sont beaucoup diversifiés, mettant en évidence notamment le polymorphisme à l'échelle de l'ADN. Il est ainsi possible de mettre en évidence, lorsqu'elles existent, les liaisons entre certains marqueurs moléculaires et des caractères d'intérêt zootechnique, appréhendés jusqu'à maintenant par leur simple mesure sur les animaux (cf. GQ, chapitre VI). Ainsi, un ensemble de circonstances ont rendu nécessaire l'approche rationnelle de l'utilisation des caractères à déterminisme génétique simple et des gènes ou marqueurs identifiés, en élevage pour eux-mêmes ou comme aides à la sélection.

B. Le polymorphisme outil de caractérisation génétique

1. La caractérisation des individus

L'identification des animaux est à la base de toute action technique collective en élevage, de nature sanitaire (prophylaxie), génétique (sélection) ou de santé publique (traçabilité). Il est par ailleurs souvent nécessaire, ou tout du moins utile, que l'on connaisse la généalogie des animaux (c'est typiquement le cas en amélioration génétique), qui est établie sur la base de déclarations. La législation française prévoit le contrôle obligatoire de filiation dans certains cas, notamment pour les reproducteurs d'insémination artificielle et les étalons de monte publique. Pour cela, on met à profit la connaissance de systèmes génétiques simples et faciles d'observation, pour comparer les génotypes d'un animal donné et ceux de ses parents supposés (cf. chapitre précédent). Les marqueurs microsatellites sont aujourd'hui employés pour cela et sont particulièrement intéressants du fait de leur grand polymorphisme : en effet, plus le polymorphisme des marqueurs observés est élevé, plus s'accroît la probabilité de détecter les fausses filiation (fausses pour cause d'erreur ou ... de fraude).

2. La caractérisation des populations⁷

Le polymorphisme peut être utilisé pour caractériser la variabilité qui existe au sein d'une population ou entre populations. Dans cette perspective, les marqueurs moléculaires sont des outils de choix, du fait de leur répartition sur l'ensemble du génome, de leur polymorphisme, de leur apparente neutralité vis-à-vis de la sélection et de l'automatisation des typages.

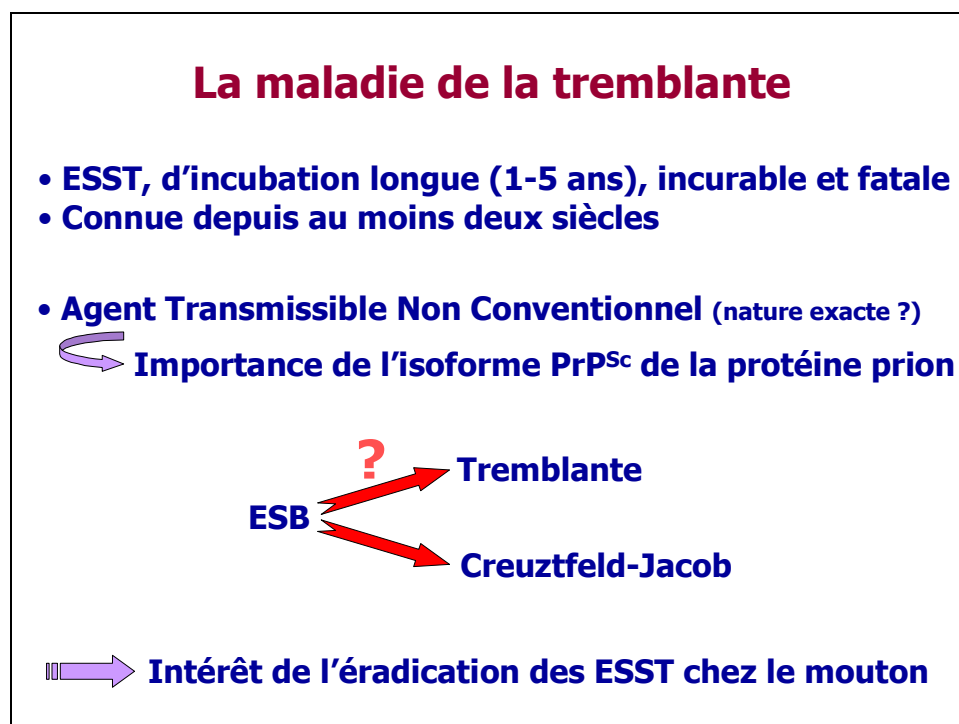
L'information de base est constituée par les fréquences, alléliques ou génotypiques, aux locus considérés, estimées sur des échantillons d'animaux des populations considérées. De nombreux indicateurs déduits de ces fréquences permettent d'apprécier comment la variabilité génétique se structure à l'intérieur des et entre les populations et d'estimer les distances la plus ou moins grande proximité génétique entre différentes populations (races, lignées, etc.) d'une même espèce.

⁷ Ces notions sont développées dans les enseignements consacrés aux ressources génétiques

C. Un exemple détaillé de gène à effet majeur : le gène "PrP" de résistance/sensibilité à la maladie de la tremblante chez le mouton⁸

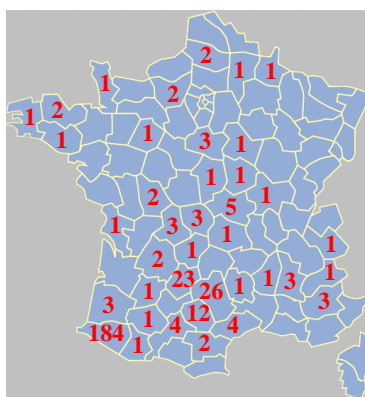
1. La maladie de la tremblante du mouton

La maladie de la tremblante (en anglais *scrapie*) est connue depuis plus de deux siècles chez le mouton. C'est une maladie neurodégénérative, sans traitement connu et d'issue fatale, du groupe des Encéphalopathies Subaiguës Spongiformes Transmissibles (ESST, dont également l'Encéphalopathie Spongiforme Bovine, ESB, et la maladie humaine de Creutzfeld-Jacob). La tremblante est réputée non contagieuse pour l'homme. Il est en revanche admis que l'ESB est transmissible à l'homme et il a été montré expérimentalement que les ovins peuvent être contaminés par l'ESB. Ainsi, tant du point de vue de la santé des cheptels ovins que des risques alimentaires liés à l'ESB, il est nécessaire d'éradiquer la maladie de la tremblante.



⁸ Du fait de l'importance de cette maladie, le gène *PrP* a fait l'objet de nombreuses publications. L'ensemble du texte de cette section reprend largement plusieurs travaux et synthèses du Département de Génétique Animale de l'INRA. Pour plus de détails, voir notamment : Elsen et al. (1997) INRA Productions Animales 10, 133-140 ; Palhière et al. (2002) Rencontres Recherches Ruminants 9, 3-9 ; Brochard et al. (2006) Rencontres Recherches Ruminants 13, 231-234.

Localisation de la tremblante en France



Nombre d'élevages atteints entre le 14 juin 1996 et le 1er août 2002

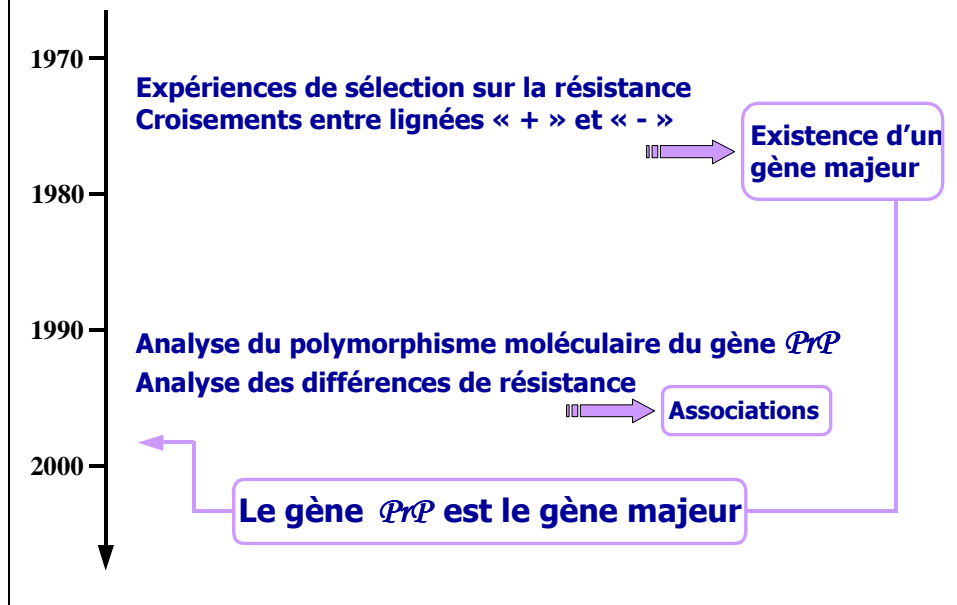
Source : AFSSA

La tremblante est présente dans tous les pays d'élevage ovin important et dotés d'un réseau d'épidémio-surveillance, à l'exception notable de l'Australie et de la Nouvelle Zélande. En France, entre juin 1996 et août 2002, 308 élevages ont été déclarés atteints de la tremblante (carte ci-dessus). Les départements du Sud, à fort cheptel ovin, ont été très touchés, particulièrement celui des Pyrénées Atlantiques (60% des troupeaux déclarés atteints).

2. Facteurs génétiques de la résistance à la tremblante

Dans les années 1970, il a été possible de réaliser des expériences de sélection sur la réponse à une injection d'agents de transmission de la maladie. La réponse observée, très importante, a permis de démontrer les bases héréditaires de la sensibilité à la maladie. A la fin des années 1970, par des croisements appropriés, l'existence d'un gène à effet majeur a été démontrée. Compte tenu du rôle de la protéine prion dans la transmission de la maladie, le gène codant pour cette protéine (gène *PrP*) a fait l'objet de nombreuses investigations. Chez l'homme et la souris, il a été montré que le polymorphisme à ce gène est associé à des différences d'incidence des ESST. Chez le mouton, de telles associations ont été mises en évidence dans les années 1990 chez plusieurs races. Aujourd'hui, il est prouvé que le gène *PrP* est bien le gène majeur mis en évidence expérimentalement vingt ans plus tôt.

Bases génétiques de la résistance à la tremblante



Le gène *PrP* et la résistance à la tremblante

Gène *PrP* → protéine PrP
 Gène très conservé chez les mammifères
 Partie codante = 1 exon, de 253 à 269 codons selon l'espèce

codon			Allèle	Effet associé
136	154	171		
Ala (A)	Arg (R)	Gln (Q)	<i>ARQ</i>	Sensibilité
Val (V)	Arg (R)	Gln (Q)	<i>VRQ</i>	Hyper sensibilité
Ala (A)	his (H)	Gln (Q)	<i>AHQ</i>	Résistance
Ala (A)	Arg (R)	Arg (R)	<i>ARR</i>	Hyper résistance

Trois codons du gène *PrP* sont impliqués (les n°s 136, 154 et 171). Les variations de résistance sont à analyser selon le polymorphisme sur ces 3 codons considérés simultanément. Les allèles sont dénommés par 3 lettres désignant les acides aminés correspondant aux 3 codons. Ainsi, l'allèle *ARQ* signifie alanine (A) en 136, Arginine (R) en 154 et glutamine (Q) en 171. Il est vraisemblable que cet allèle soit l'allèle ancestral, dont ont dérivé par mutation ponctuelle les autres allèles connus, parmi lesquels 3 principaux : *VRQ*, *AHQ* et *ARR*.

Plusieurs résultats expérimentaux ont permis de préciser les effets de ces différents allèles sur la sensibilité des ovins à tremblante. Il en ressort que les allèles *VRQ* (systématiquement) et *ARQ* (en général) sont associés à une sensibilité à cette maladie, alors que les allèles *AHQ* (en général) et *ARR* (systématiquement) sont associés à une résistance. Les allèles de résistance (*AHQ* et *ARR*) sont partiellement dominants sur les autres allèles, y compris sur *VRQ* qui, lui-même, est partiellement dominant vis-à-vis des autres allèles (*ARQ* et des allèles plus rares).

3. Bilan des différentes populations ovines vis-à-vis des allèles au gène *PrP*

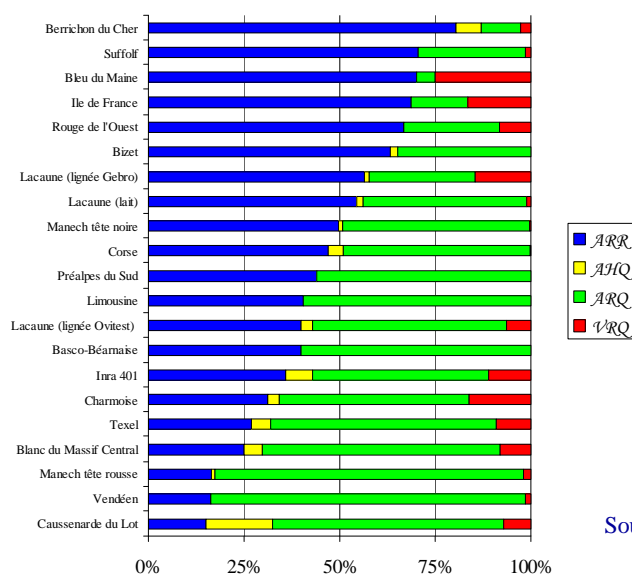
Les techniques de biologie moléculaire permettent le typage sans ambiguïté des animaux pour le gène *PrP*. En France, 29 races⁹ ont été examinées, sur des échantillons d'une centaine d'animaux par race (figure ci-après). Globalement, les allèles *ARQ* et *ARR* sont les plus courants, alors qu'*AHQ* est rare. La situation la plus favorable est celle du Berrichon du Cher, avec une fréquence d'*ARR* supérieure à ¾. A l'opposé, plusieurs races ont une fréquence d'*ARR* inférieure à ¼, dont la Manech Tête Rousse, élevée dans les Pyrénées Atlantiques, département signalé comme le plus touché par la maladie (cf. plus haut).

4. Questions soulevés par une éventuelle sélection sur le gène *PrP*

Compte tenu de l'intérêt d'éradiquer les ESST chez le mouton, des effets du gène *PrP* et de l'existence d'un protocole de typage moléculaire pour ce gène, il est tentant d'effectuer dans les populations ovines une sélection en faveur de l'allèle *ARR*, dans le but de rendre le cheptel ovin entièrement résistant à la tremblante. Avant d'entreprendre une telle sélection, il est néanmoins nécessaire d'évaluer certains risques et de prendre les moyens de les minimiser.

⁹ Pour diverses informations sur l'ensemble des races ovines françaises, consulter : <http://www.inapg.fr/dsa/especes/ovins.htm>

Fréquences alléliques au gène *PRP* dans les principales races ovines françaises (état 1999-2000)



Source : Pallière et al (INRA), 2002

Sélection en faveur d'*ARR* : questions soulevées

- **Universalité de la résistance ?**
- **Vers des populations de porteurs sains ?**
- **Conséquences négatives pour d'autres caractères ?**
- **Coûts et précautions de mise en œuvre ?**

a. L'universalité de la résistance

Outre l'agent de l'ESB, il existe plusieurs souches d'agents infectieux de la tremblante que l'on sait distinguer par bio-essais sur lignées de souris sensibles. Il faut donc s'assurer que la résistance conférée par l'allèle ARR est valable pour toutes les souches (absence d'interaction entre le génotype de l'hôte, le mouton, et l'agent pathogène). En ce qui concerne les agents naturels de transmission de la tremblante, la convergence des résultats obtenus avec des races variées, élevées dans des milieux variés, plaide pour l'universalité de la résistance, étant donné qu'ils ne mentionnent aucun ovin ARR/ARR atteint de tremblante. En ce qui concerne l'agent de transmission de l'ESB, aucun homozygote ARR/ARR inoculé par voie orale n'a été atteint par la maladie. Un seul cas d'homozygote ARR/ARR atteint après une inoculation intra-cérébrale massive a été publié (à grands bruit, en novembre 2002). Il faut néanmoins relativiser l'annonce de ce dernier cas, compte tenu du mode d'inoculation intra cérébrale, non pertinent pour mimer ce qui se passe dans les élevages et qui pose de gros problèmes d'interprétation biologique (il est ainsi possible d'infecter par cette voie des porcs, espèce non sensible naturellement aux ESST). On peut souligner, par ailleurs, que des expériences ont mis en évidence la très grande difficulté de la conversion *in vitro* de la protéine prion codée par des allèles ARR vers l'isoforme PrP^{Sc} spécifique de la tremblante.

b. Résistance et portage de l'agent infectieux

Comme pour toute maladie, il faut s'interroger sur la possibilité pour que les individus résistants soient des porteurs sains de l'agent infectieux, la résistance pouvant consister, par exemple, en un simple allongement de la durée d'incubation. Ainsi, les animaux résistants pourraient constituer des réservoirs d'agents pathogènes, situation défavorable à l'éradication de la maladie. Cette question a été abordée expérimentalement par l'analyse de l'accumulation de l'isoforme PrP^{Sc} dans les tissus lymphoïdes ou nerveux, les amygdales en particulier. Il en ressort que, tant en situation de contamination naturelle qu'après inoculation de l'agent de l'ESB, les homozygotes ARR/ARR ne sont pas porteurs, même après un délai de 18 mois, alors que les homozygotes VRQ/VRQ le sont dès le 1er ou le 2ème mois suivant la contamination. Ces résultats sont confirmés par des analyses effectuées sur des animaux d'élevages.

c. Liaison génétique entre résistance et autres caractéristiques

Il est nécessaire de vérifier que cette nouvelle sélection n'est pas susceptible de détériorer la moyenne des populations pour des caractères, déjà sélectionnés ou non. Dans le cas présent, cela serait possible si l'allèle *ARR* était associé, de façon fortuite ou causale, à des valeurs génétiques inférieures à la moyenne pour certains caractères d'intérêt. Les analyses effectuées à ce jour, sur des effectifs importants d'animaux, n'ont pas permis de détecter de telles associations, qu'il s'agisse de caractères de reproduction (fertilité, taille de portée), de production laitière ou de croissance.

d. Coûts et précautions de mise en oeuvre

Il est nécessaire d'appréhender les coûts et les risques, autres que ceux soulignés plus haut, induits par la mise en oeuvre d'une sélection sur le gène *PrP*. Les coûts directs sont liés au typage moléculaire et interdisent d'envisager de typer tous les animaux d'une population donnée. Il est ainsi nécessaire d'appliquer cette mesure à certains animaux seulement, tout particulièrement ceux qui sont susceptibles d'avoir une descendance nombreuse et/ou une forte influence sur l'évolution génétique de la population : jeunes mâles candidats à la sélection, potentielles mères de béliers, etc.

Des coûts indirects sont liés à la dispersion des efforts de sélection : même en absence de liaison génétique négative, appliquer à moyens constants une pression de sélection sur un caractère qui n'était pas prise en compte auparavant induit obligatoirement une diminution du progrès génétique sur les autres caractères sélectionnés (voir chapitre V). La seule façon d'éviter ce manque à gagner est d'accroître les moyens mis globalement pour faire fonctionner le programme de sélection, notamment augmenter le nombre d'animaux contrôlés et/ou évalués, c'est-à-dire le nombre de candidats à la sélection.

Enfin, le fait de sélectionner les porteurs d'un allèle donné, voire les seuls homozygotes, est susceptible de grandement favoriser les représentants de certaines familles seulement, et de créer ainsi de forts goulets d'étranglement (cf. GP). Cela peut conduire à un rétrécissement de la base génétique des populations et à une accélération de l'accroissement de consanguinité.

Ces phénomènes ont été observés dans la race porcine Landrace Français¹⁰, suite à l'élimination au cours des années 1980 de l'allèle "sensible" au gène de sensibilité à l'halothane¹¹, gène qui a de multiples effets, notamment sur la musculature des animaux et la qualité de leur viande (voir plus loin). La mise en oeuvre d'une sélection sur le gène *PrP* nécessite l'application de méthodes permettant de combiner efficacité génétique à court terme et préservation de la variabilité génétique à moyen terme. Les précautions à prendre dans ce domaine sont d'autant plus grandes que la fréquence initiale de l'allèle *ARR* est faible.

5. Le programme de sélection sur la résistance à la tremblante mis en place en France

Plusieurs pays européens, notamment le Royaume Uni, la France et les Pays Bas, ont mis en place un programme national de lutte contre la maladie de la tremblante. Le dispositif français comprend 3 volets¹² : (i) un réseau de surveillance clinique, (ii) des mesures de police sanitaire (retrait de matériaux à risque, plans d'abattage sélectifs) et (iii) un programme de sélection sur la résistance à la tremblante. Il est en effet établi qu'une sélection en faveur de l'allèle *ARR* du gène *PrP*, peut conduire à la généralisation de la résistance des populations ovines aux ESST, *a minima* dans les conditions naturelles de contamination.

Le programme national de sélection sur la résistance à la tremblante a été défini et avalisé en 2001-2002. Il concerne toutes les races ovines, y compris celles faisant l'objet de mesures conservatoires du fait de leur faible effectif. L'objectif général de ce programme est la protection générale de l'ensemble du cheptel ovin, ce qui se décline en deux temps :

- Dans l'immédiat, élimination de l'allèle *VRQ* d'hyper sensibilité. Cet objectif est le plus facile à atteindre, du fait de la fréquence initiale généralement faible de cet allèle et de la possibilité de déterminer sans ambiguïté le génotype des animaux typés.
- A moyen terme, généraliser l'utilisation à grande échelle de béliers homozygotes *ARR/ARR*, en vue d'un double effet : d'une part, protection jusqu'à l'âge de l'abattage des agneaux issus de ces béliers et, d'autre part, diffusion de l'allèle *ARR* dans l'ensemble de la population.

La mise en oeuvre de ce programme s'appuie sur les dispositifs et les structures en place, dans le cadre de la réglementation en vigueur de la sélection animale.

¹⁰ <http://www.inapg.fr/dsa/especes/porcins/lrf.htm>

¹¹ <http://www.inapg.fr/dsa/uvf/AG/genes/haloth/haloth.htm>

¹² <http://www.agriculture.gouv.fr/alim/sant/mala/tremblante>

Le programme national de sélection sur la résistance à la tremblante

A partir de 2002

Objectif = protection de l'ensemble du cheptel

- Elimination de l'allèle *VRQ*
- Utilisation de béliers *ARR/ARR*

En s'appuyant sur les dispositifs en place
En aménageant les programmes

Utilisation
de gènes connus



Chapitre III du polycopié
<http://www.inapg.inra.fr/dsa/uvf/AG/genes/genintro.htm>

Bilan après 4 ans de sélection sur la résistance à la tremblante

- 400 000 animaux typés
- Allèles de *PrP* (estimation chez les mâles, toutes races) :
 - Quasi-disparition de *VRQ*
 - Augmentation de la fréquence d'*ARR*, + 35% (approx.)
- Autres conséquences (étude spécifique sur 4 races) :
 - Sélection de béliers qui ne l'auraient pas été auparavant
 - Effet sur les autres caractères, selon la race : stagnation des niveaux génétiques ou maintien des évolutions antérieures
 - Effet sur la variabilité génétique : néant, hormis à *PrP* lui-même et dans les zones très proches du génome

Source : Brochard et al., 2006

D. L'utilisation des caractères à déterminisme génétique simple et des gènes correspondants

L'exemple détaillé du gène de résistance à la tremblante du mouton illustre parfaitement la démarche qui doit être suivie dans une approche rationnelle de l'utilisation en élevage de caractères à déterminisme génétique simple : (i) observation du phénomène et détermination du mécanisme héréditaire en cause, (ii) inventaire des liaisons éventuelles entre le génotype recherché et d'autres caractères et (iii) mise en oeuvre des moyens susceptibles d'augmenter la fréquence de l'allèle ou du génotype recherché, voire fixation de ce génotype, et aménagement des programmes déjà en place pour aller dans ce sens. Les caractères recherchés doivent avoir un intérêt économique prouvé (ceci est d'ailleurs valable quel que soit le déterminisme génétique en cause), tout effort de sélection en leur faveur se faisant au détriment de l'effort de sélection fait sur d'autres caractères. Le tableau de la page suivante présente, uniquement à titre d'exemple et d'illustration, un panorama non exhaustif, de l'utilisation de ce type de caractère et des gènes correspondants.

E. Bilan et perspectives

Les caractères à déterminisme génétique simple ont eu une grande importance historique en élevage, et leur utilisation répond aujourd'hui essentiellement à des objectifs économiques. Il apparaît clairement que les découvertes de gènes à effets visibles ou d'autres polymorphismes discontinus sont étroitement liées aux efforts d'observation et d'analyse qui sont effectués au sein des populations animales domestiques. Il en va de même pour la détection de gènes induisant des variations sur des caractères continus : gènes majeurs, détectables par analyse statistique de performances et de généalogies, ou QTL, à effet moins fort et détectables avec les mêmes informations plus des marqueurs moléculaires (voir GQ chapitre VI).

Tableau 1. Panorama non exhaustif de l'utilisation en élevage et en sélection des caractères à déterminisme simple et des gènes correspondants

Famille de caractères	Objectif recherché	Espèce / production / race	Caractère / gène	Usage / Action
Coloration, aspect du poil ou du plumage	Marque de fabrique	Poule pondeuse	Couleur des œufs	Répondre aux habitudes des consommateurs
		Poulet de chair sous label	Gène « cou nu »	Distinguer le produit des autres
		Poulet de Bresse AOC	Couleur des pattes	Distinguer le produit des autres
		Bovins laitiers en croisement avec une race bouchère ; Bovins allaitants en ranching	Couleur de la robe	Distinguer les animaux croisés de ceux de race pure
		Bovins laitiers en croisement avec la race Holstein (par ex.)	Couleur de la robe	Procéder à des infusions de gènes extérieurs tout en gardant l'aspect extérieur
	Adaptation au milieu	Poule	Gène « cou nu »	Production en conditions chaudes
		Bovins (ex. race Abondance)	Tâche colorée autour de l'œil	Protection vis-à-vis du cancer de l'œil
Intérêt économique direct	Animaux producteurs de laine ou poil	Couleur et texture de la fibre	Valeur marchande de la fibre	
	Poule	Gène de coloration lié au sexe	Détermination aisée du sexe des poussins à la naissance	
Cornage	Minimisation de la gène et des risques	Caprins laitiers	Gène « motte »	Impossibilité de sélectionner l'allèle « sans corne » car celui-ci induit aussi de graves troubles de la fertilité (gène pléiotrope).
		Bovins	Gène « sans corne »	Recherches en cours pour l'identification du gène
Anomalies héréditaires graves	Eradiquer l'anomalie	Bovins (notamment race Holstein)	Gènes BLAD (déficience immunitaire), « Bull-dog » (malformations), ...	Observatoire national des anomalies (2002). Recherche du gène responsable et détection, par typage, des hétérozygotes porteurs sains.
Anomalies de taille	Réduction des coûts d'élevage	Poule	Gène de nanisme, lié au sexe, avec l'allèle « nain » récessif	Introduction dans les lignées maternelles employées dans les plans de croisement.
Développement musculaire et qualité de la viande	Accroissement du taux de muscle dans la carcasse	Porc	Gène « Halothane » : l'allèle « n » induit une hypertrophie musculaire mais dégrade la qualité de la viande	Fixation de l'allèle « n » chez la race Piétrain (Belgique, Allemagne) ; élimination dans les lignées maternelles ; obtention de produits croisés hétérozygotes.
		Porc	Gène « RN »	Élimination de l'allèle responsable d'une mauvaise qualité de viande
		Bovins	Gène de la myostatine (caractère « culard »)	Fixation de l'allèle « culard » chez les races bovines Blanc Bleu Belge et Piémontaise, la race ovine Texel Belge ; élimination dans les races laitières et dans les races allaitantes en milieu contraignant.
		Ovins		
Protéines du lait	Amélioration de la qualité fromagère des laits	Bovins	Gène de la κ -caséine	Typage des taureaux d'insémination (depuis 1988)
		Caprins	Gène de la caséine α -s ₁	Typage des boucs d'insémination

Depuis la fin des années 1990, les résultats de programmes importants de détection de QTL s'accumulent : des QTL sont détectés pour quasiment tous les caractères mesurés ; pour un caractère donné, le(s) QTL détecté(s) explique(nt) un total de 20 et 40% de la variance génétique. Une première conséquence est l'emploi en sélection de ces informations. Par exemple, typer des animaux pour les marqueurs proches des QTL peut permettre des gains de précision (plus d'information pour l'évaluation génétique) ou des gains de temps (évaluation génétique précoce, avant même de réaliser la mesure des phénotypes habituellement nécessaires). Ainsi, depuis 2001 en France, la Sélection Assistée par Marqueurs (SAM) est en oeuvre chez les 3 principales races bovines laitières. Ici, l'information moléculaire ne se substitue pas aux autres sources d'information mais est intégrée dans un dispositif d'évaluation génétique qui combine toutes les informations disponibles.

Une autre conséquence est la possibilité d'analyser finement le génome, tout d'abord dans les régions chromosomiques où ont été détectées les QTL. Des techniques de cartographie fine et de cartographie comparée (entre espèces) permettent ainsi de mettre en évidence, dans le gène, la mutation responsable des variations observées. Cela ouvre les possibilités de Sélection Assistée par le Génotypage (SAG), plus efficace encore que la SAM puisque l'on joue directement sur les gènes impliqués et non sur des marqueurs à proximité.

Le développement des outils et méthodes de la génomique fonctionnelle (transcriptome, protéome, ...) ouvrent la voie à une analyse à grande échelle du fonctionnement du génome. Des programmes de recherche se mettent en place dans ce sens, dont en France depuis 2002, le programme AGENAE (Analyse du GENome des Animaux d'Élevage) qui concerne quatre espèces : bovins laitiers, porcs, poulet et truite. Ce programme multidisciplinaire, sous l'égide de l'INRA, bénéficie du soutien financier de plusieurs organismes professionnels d'élevage et de sélection. Il se développe autour de plusieurs axes : développement des outils de la génomique, analyse fonctionnelle (comparaison de l'expression des gènes à différents stades physiologiques, chez des animaux de statuts infectieux différents, etc.), génétique (comparaison de génotypes, diversité génétique), bio-informatique, et infrastructures (Centres de Ressources Biologiques). Ce type de programme, au long cours, doit apporter de nouvelles connaissances sur les mécanismes fins de fonctionnement du génome. Il va fournir également de nouveaux phénotypes, l'expression des gènes, dont la variabilité pourra être analysée (comme celle d'autres caractères) et sans doute exploitée en sélection.

Enfin, on est capable de cloner un fragment d'ADN, d'associer un gène de structure à des promoteurs et autres éléments régulateurs, et d'implanter la "construction" ainsi obtenue dans le génome d'un organisme vivant. Appliquée précocement (stade une cellule, par exemple), cette technique de génie génétique et de transfert de gènes permet de modifier directement le génome des êtres vivants. Les premiers résultats ont été observés en 1982, des souris géantes ayant été obtenues par transfert du gène de l'hormone de croissance du rat. Des applications, avec des résultats beaucoup moins spectaculaires, ont déjà été effectuées chez des animaux de ferme. Certes, beaucoup d'efforts restent à faire en vue de maîtriser les effets secondaires indésirables du transfert de gènes, notamment en ce qui concerne la viabilité et la fertilité des animaux ainsi modifiés. Des applications en milieu très maîtrisé et confiné ont déjà cours, notamment la production de molécules d'usage pharmaceutique dans le lait de femelles transgéniques (le gène de structure de la molécule ayant été associé avec le régulateur d'un gène de lactoprotéine). Mais il n'est pas du tout certain que notre société accepte l'usage d'animaux transgéniques à des fins alimentaires : il suffit de voir la sensibilité du grand public et/ou des médias lors des crises sanitaires qu'ont traversées depuis le milieu des années 1990 les filières animales et la mobilisation en Europe contre l'emploi des plantes transgéniques (OGM). A l'évidence, beaucoup de réflexion et de dialogue ouvert sont encore nécessaires afin d'apporter des réponses aux questions éthiques et juridiques que pose cette faculté d'agir directement sur le génome des animaux.

IV. L'ÉVALUATION GÉNÉTIQUE DES REPRODUCTEURS

Lorsqu'il choisit des reproducteurs, le sélectionneur attache évidemment de l'importance aux aptitudes propres de ces animaux : ceux-ci doivent notamment être aptes à se reproduire avec efficacité. Mais il cherche surtout à prédire la valeur moyenne de la descendance engendrée par ces reproducteurs. Cette valeur de la descendance est fonction, d'une part, des effets moyens des gènes reçus des parents, et donc de leur valeur génétique additive et, d'autre part, des effets d'interaction entre gènes paternels et maternels (voir GQ). L'évaluation génétique consiste ainsi à prédire¹³ la valeur génétique additive des animaux (la prise en compte des effets d'interaction sera vue au chapitre VI). Les valeurs génétiques¹⁴ estimées, désignées par le terme d'index de valeur génétique) servent à classer les candidats à la sélection. L'évaluation génétique est donc une phase essentielle des programmes de sélection des animaux.. Dans tout ce chapitre, nous traitons de *l'évaluation génétique pour un seul caractère* (nous évoquerons la sélection multicaractères au chapitre V).

Bases du choix d'un futur reproducteur

A/ ~~Pour ses performances~~

B/ Pour la valeur attendue de sa descendance

C/ ~~Pour ses beaux yeux~~

→ Comment classer les candidats à la sélection sur la valeur de leur descendance future ?

Sur la base de leur valeur génétique additive

¹³ En statistiques, on dit que l'on « prédit » une valeur génétique. Par abus de langage, on dit souvent « estimer ».

¹⁴ Dans la pratique de l'amélioration des animaux, « valeur génétique » signifie implicitement « valeur génétique additive ». C'est cette convention que nous emploierons tout au long de ce document.

Notion de valeur génétique additive

(Rappels)

$$P = \mu + \textcircled{G} + E \quad \longrightarrow \quad G = \boxed{A} + \boxed{D}$$

Somme des effets moyens des gènes

$$E(A_i | A_p) = 1/2 A_p$$

$$E(P_i | A_p) = \mu + 1/2 A_p$$

Effets d'interaction
entre les gènes
dépendent de la manière
dont les couples sont constitués

→ Cours de génétique quantitative

Evaluation génétique

Consiste à prédire (estimer)
la valeur génétique (additive)
d'un animal ou d'un ensemble d'animaux

Indexation

- sert à classer les candidats à la sélection -

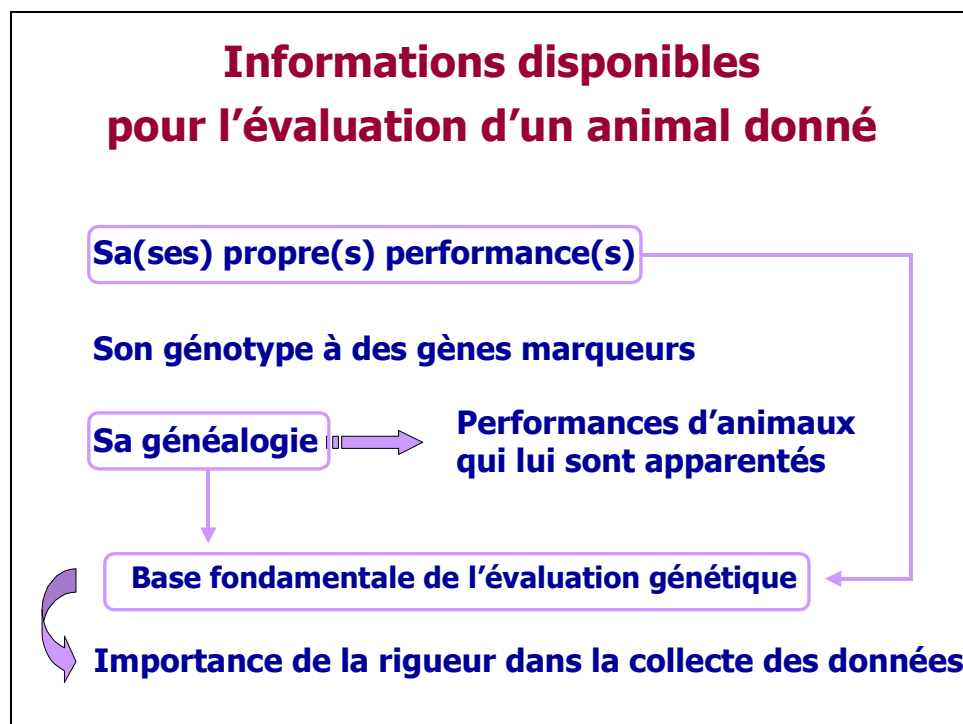
Index de valeur génétique

Estimated Breeding Values

A. Les informations disponibles pour l'évaluation génétique

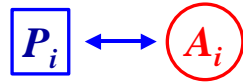
1. Comment classer des candidats sur la base de la valeur de leur descendance future ?

Un moyen de comparer les candidats pour ce qu'ils peuvent transmettre à leur descendance est de faire procréer à chacun un grand nombre de descendants et de les classer sur la valeur moyenne de leur descendance. Nous verrons d'ailleurs que ce protocole est à la base d'une des méthodes possibles de sélection. Cependant, cette procédure est compliquée et coûteuse à mettre en oeuvre, si bien qu'elle n'est pratiquement applicable qu'à un petit nombre de candidats ; de plus, les résultats sont, par nature, longs à obtenir. Aussi est-il nécessaire de disposer d'une méthode générale, qui permette de se faire une idée de la descendance future d'un candidat même en absence de descendants déjà connus, et qui valorise toutes les informations dont on dispose à son sujet. Ces informations sont de différentes natures : des performances (valeurs phénotypiques), le génotype à des gènes connus ou à des marqueurs (cf. chapitre précédent), et des généalogies. C'est dire l'importance cruciale en génétique animale de pouvoir disposer de protocoles rigoureux de contrôle des performances des animaux et d'un enregistrement fiable de leur état civil.



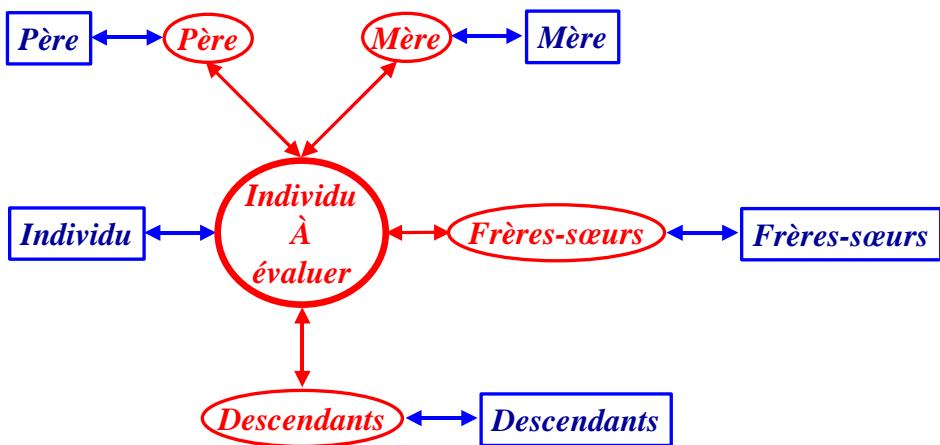
En quoi les performances sont-elles utiles ?

Approche statistique : La valeur génétique est une variable aléatoire, que l'on cherche à prédire



Car A_i est contenue dans P_i

En quoi les performances sont-elles utiles ?



Du fait de l'apparentement

2. Principales informations utilisées pour l'évaluation génétique

Pour prédire la valeur génétique d'un animal, le plus simple est de mesurer sa performance : il s'agit d'une *évaluation sur propre performance*. Si le choix des reproducteurs ne s'effectue que sur cette information, on parle de sélection individuelle ou *sélection massale*. Cette procédure peut cependant manquer d'efficacité et, surtout, n'est pas toujours applicable (cas de caractères ne s'exprimant que dans un seul sexe, production laitière ou d'œufs, caractères dont la mesure nécessite l'abattage de l'animal, comme le taux de muscle dans une carcasse).

Nous savons cependant que des animaux apparentés entre eux ont des gènes en commun et que leurs valeurs génétiques sont corrélées (voir GQ, chapitre IV). Il est donc possible d'utiliser des performances mesurées sur des individus apparentés au candidat à la sélection pour prédire la valeur génétique de ce dernier. Il s'agit alors d'une *évaluation sur apparentés*, la méthode de sélection correspondante étant désignée généralement sous le terme de *sélection généalogique*. Pour que cette procédure soit efficace, il est toutefois nécessaire que le lien de parenté soit relativement proche. Aussi distingue-t-on principalement trois méthodes selon les performances d'apparentés qui sont utilisées :

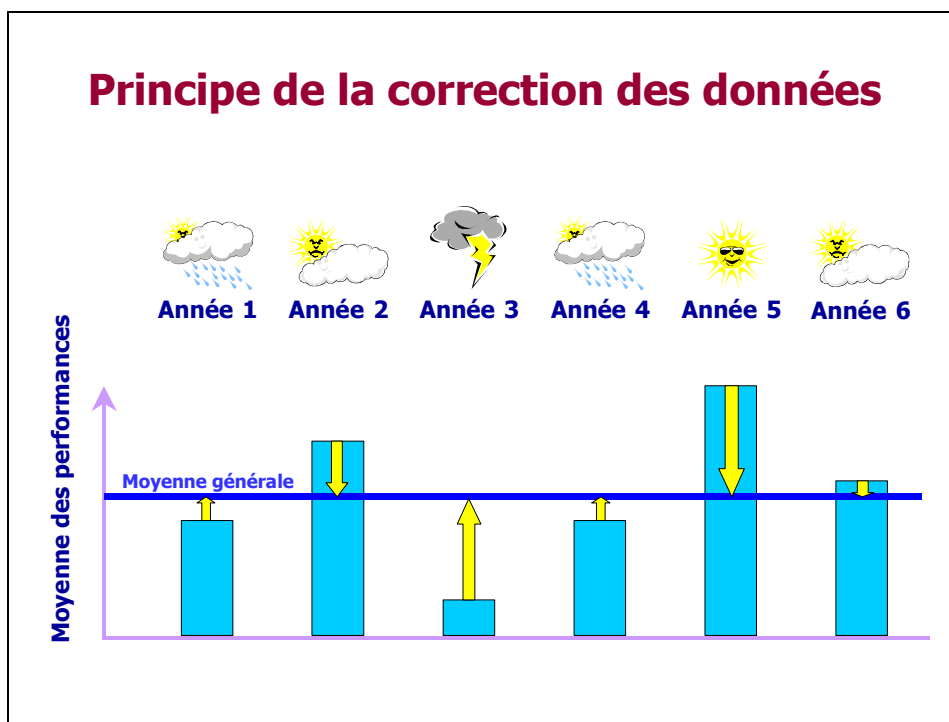
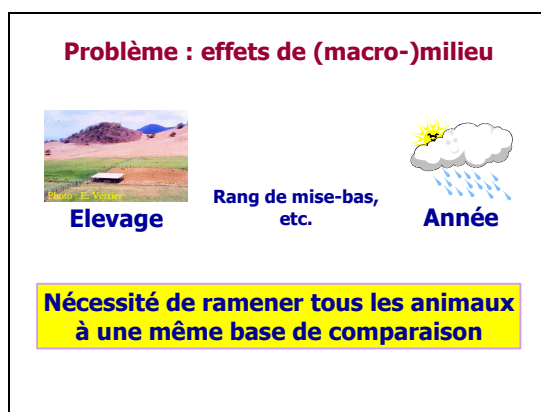
- Performances des parents du candidat : évaluation, ou sélection, *sur ascendance*
- Performances des descendants du candidat : évaluation, ou sélection, *sur descendance*
- Performances des frères et sœurs du candidat : évaluation, ou sélection, *sur collatéraux*

Dans le présent document, nous envisageons ces différentes possibilités essentiellement de façon séparée. Dans la pratique, bien sûr, on est le plus souvent amené à utiliser l'ensemble des informations disponibles pour un candidat à un instant donné : on parle alors d'*évaluation*, ou de sélection, *combinée*.

3. Nécessité de corriger les performances pour les effets de milieu

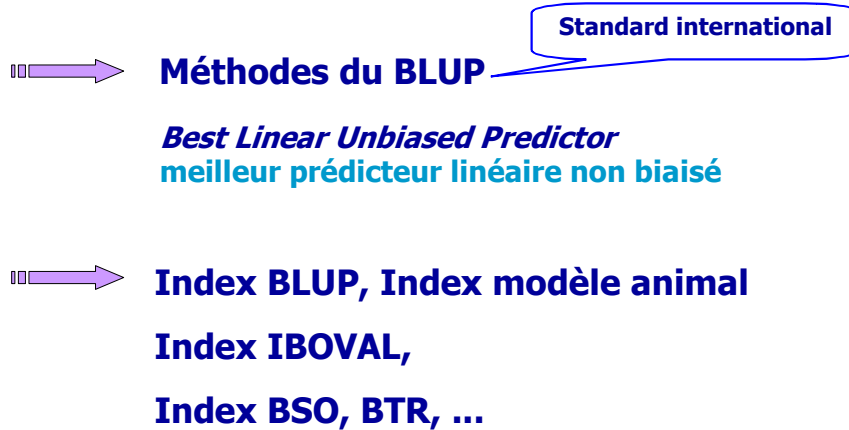
Il existe des facteurs de milieu qui ont des effets systématiques sur les performances des animaux : élevage où sont les animaux, années où la performance a été réalisée, etc. Ces facteurs constituent le « macro-milieu » (voir GQ, chapitre II). Une des conditions de la fiabilité des index de valeur génétique est que les performances soient corrigées pour les effets des facteurs de macro-milieu. Dans la pratique, cela signifie tout d'abord que ces facteurs soient enregistrés (il faut savoir dans quelles conditions ont été placés les animaux). Ensuite, pour un facteur particulier, il faut estimer l'effet de chacune des modalités et

retrancher à chaque performance la valeur de l'effet correspondant à la modalité dans laquelle cette performance a été réalisée. Il s'agit cependant d'une vision simplifiée des choses et on doit avoir recours à des méthodes statistiques élaborées pour traiter correctement ce problème. A ce jour, la méthode la plus fiable est la méthode du BLUP (Best Linear Unbiased Predictor), développée par Henderson (Université de Cornell, USA), notamment dans une version particulièrement efficace dite du « modèle animal ». Dans tout ce qui suit, *nous raisonnons sur des performances correctement corrigées pour les effets de macro-milieu.*



Correction pour les effets de milieu

Problème moins simple qu'il n'en a l'air !



B. Le calcul des index de valeur génétique

1. Rappels sur la notion de meilleur prédicteur (voir l'enseignement de statistiques)

Soit deux variables de distributions quelconques, X et Y , la première étant la variable observée (ou variable prédictrice) et la seconde étant la variable à prédire. La relation entre le prédicteur que l'on recherche (\hat{Y}) et la valeur vraie est donnée par :

$$Y = \hat{Y} + e$$

où e représente une erreur résiduelle. En statistiques, le « meilleur prédicteur » est celui qui minimise l'espérance du carré de l'erreur (méthode des moindres carrés). *Dans tous les cas*, le meilleur prédicteur est l'espérance de Y sachant X (espérance conditionnelle) :

$$\hat{Y} = E(Y|X)$$

On montre que, par construction, d'une part, l'erreur résiduelle est d'espérance nulle et donc sa variance est minimale aussi et, d'autre part, cette erreur est non corrélée à la valeur prédite.

Quand X et Y ont une distribution conjointe normale, les espérances conditionnelles de Y s'ajustent sur une droite : le meilleur prédicteur de Y est obtenu par régression linéaire. L'équation de la droite de régression est donnée par l'expression suivante :

$$\hat{Y} = E(Y|X) = E(Y) + \frac{\text{cov}(X, Y)}{\text{var}(X)} [X - E(X)]$$

Cette formulation est tout à fait équivalente à celle plus classiquement employée, à savoir :

$$\hat{Y} = a + bX, \quad \text{avec } b = \frac{\text{cov}(X, Y)}{\text{var}(X)} \quad \text{et } a = E(Y) - bE(X)$$

Le calcul des index de valeur génétique s'effectue selon ce principe, dans le cadre de la normalité des distributions et d'une liaison linéaire entre les variables considérées.

2. Un exemple simple : l'évaluation à partir de la performance individuelle

Soit P_i la performance mesurée sur l'animal i et A_i sa valeur génétique à prédire. En appliquant au cas présent ce qui a été vu au § 1, l'index de valeur génétique s'écrit :

$$\text{Index} = \hat{A}_i = E(A_i|P_i) = b(P_i - \mu)$$

μ désignant la moyenne des performances dans la population. La pente de la droite (b) est le coefficient de régression de la valeur génétique (A_i) sur la performance (P_i) :

$$b = b_{A_i/P_i} = \text{cov}(A_i, P_i) / \text{var}(P_i)$$

Nous raisonnons dans le cadre d'un modèle (voir GQ) où la performance est fonction de la valeur génétique (A_i), la valeur de dominance (D_i) et la résiduelle environnementale (E_i) :

$$P_i = \mu + A_i + D_i + E_i$$

Par construction, les 3 variables A , D et E sont non corrélées. On montre alors facilement que le coefficient de régression (b) est égal au coefficient d'héritabilité (h^2) du caractère :

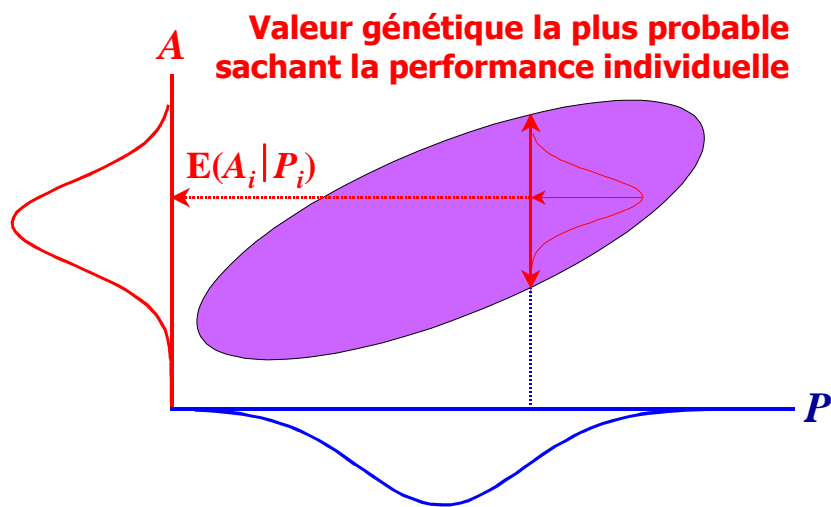
$$b = \text{cov}(A_i, A_i + D_i + E_i) / \text{var}(P_i) = \sigma_A^2 / \sigma_P^2 = h^2$$

En définitive, l'index **établi à partir de la performance individuelle** s'écrit :

$$\boxed{\hat{A}_i = \text{Index sur performance individuelle} = E(A_i|P_i) = h^2(P_i - \mu)}$$

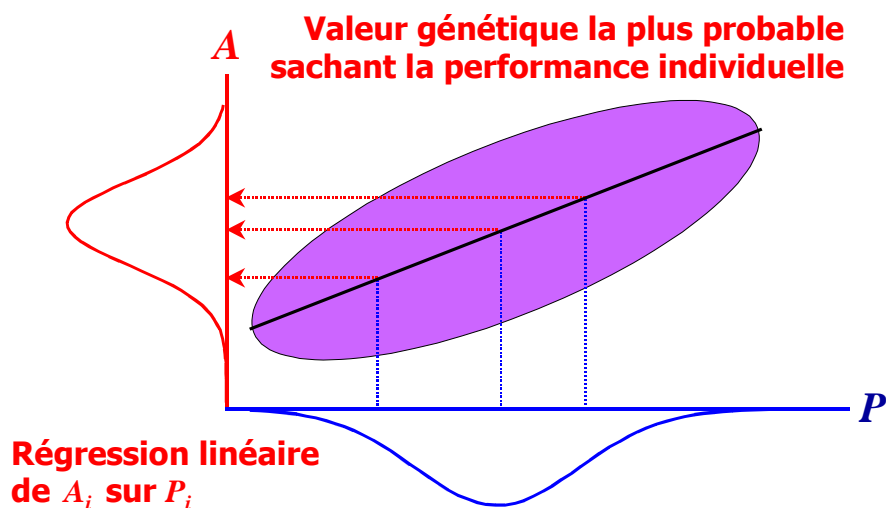
Un exemple simple

Evaluation sur la performance individuelle



Un exemple simple

Evaluation sur la performance individuelle



Index : définition

Index de valeur génétique

=

**Valeur génétique la plus probable
compte tenu de la performance**

=

**Espérance de la valeur génétique sachant la
performance**

=

$$b \times (P - \mu)$$

Ecriture de l'index de valeur génétique

Cas de la performance individuelle

$$\text{Index} = h^2 (P_i - \mu)$$

**Rappel : h^2 = part des différences entre individus
qui est d'origine génétique additive**

3. La prise en compte des apparentés

Nous savons (cf. GQ, chapitre IV) que les valeurs génétiques de deux individus apparentés sont corrélées. En effet, ayant au moins un ancêtre commun, les apparentés partagent des gènes en commun, ce qui induit une corrélation entre leurs valeurs génétiques. La performance d'un animal apparenté à l'animal à évaluer est donc, par ce jeu de corrélations, informative vis-à-vis de la valeur génétique que l'on cherche à estimer.

Le principe est le même que précédemment. Soit P_j la performance d'un animal j , apparenté à l'animal i que l'on cherche à évaluer, nous avons :

$$\text{Index} = \hat{A}_i = E(A_i | P_j) = b_{A_i/P_j} (P_j - \mu)$$

avec,

$$b_{A_i/P_j} = \text{cov}(A_i, P_j) / \text{var}(P_j)$$

L'expression de la covariance entre A_i et P_j nous est donné *a priori* par la formule de covariance entre apparentés (GQ, chapitre IV) :

$$\boxed{\text{cov}(A_i, P_j) = 2\Phi_{ij}\sigma_A^2}$$

Le coefficient de parenté est une proportion, qui vaut $1/4$ dans le cas parent-descendant, $1/4$ pour des pleins frères-sœurs, $1/8$ pour des demi frères-sœurs, etc¹⁵. On voit donc que la covariance entre la valeur génétique de l'animal à évaluer (A_i) et la performance d'un seul apparenté (P_j) est plus faible que la covariance entre A_i et la performance individuelle (P_i). Outre ce que nous dit l'équation ($2\Phi_{ij} < 1$), ceci est logique : le lien avec la valeur génétique individuelle est plus fort pour la performance individuelle que pour celle d'un apparenté.

Dans la pratique, on raisonne sur la moyenne des performances d'animaux ayant le même lien de parenté avec l'animal à évaluer : moyenne des frères-sœurs, des descendants, etc. Dans ce cas, l'expression de la variance de la prédictrice [$\text{var}(P_j)$] doit tenir compte des liaisons entre toutes les performances entrant dans la moyenne. Cette corrélation est donnée *a priori*, là encore, par la formule de la covariance entre apparentés. Le poids accordé à la moyenne des

¹⁵ <http://www.inapg.fr/dsa/uvf/GP/Phi/Phintro.htm>

performances d'apparentés de même type dépend de l'héritabilité du caractère considéré et du nombre total d'apparentés considérés (Tableau 2). Il n'est pas nécessaire, à ce stade, de retenir par cœur ces expressions, qui sont données à titre d'illustration. La diapositive qui suit n'en est qu'une schématisation. On pourra retenir les propriétés suivantes :

- Le poids accordé à une performance individuelle est toujours égal à l'héritabilité du caractère : c'est un point de référence.
- Le poids accordé à la moyenne des performances individuelles est plus élevé que celle accordée à une seule (on a plus d'information).
- Le poids accordé à la performance d'un ancêtre est toujours plus faible que l'héritabilité et décroît avec le rang d'ascendance.
- Le poids accordé à la moyenne des descendants dépasse la valeur de l'héritabilité à partir d'un nombre de descendants d'autant plus petit que l'héritabilité est faible.

Schématisation de l'écriture de l'index

$$\text{Index} = c \times f(n, h^2) \times (P - \mu)$$

↙ *Dépend du décalage de génération par rapport au candidat à évaluer*
→ *Poids accordé à la performance ou à la moyenne des n performances utilisées*

- Performance d'un parent direct → $c = 1/2$
- Performance individuelle ou d'un collatéral → $c = 1$
- Performance d'un descendant direct → $c = 2$

Tableau 2. Expression des index de valeur génétique fondés sur les performances individuelles ou sur celles d'un certain type d'apparentés.

$$h^2 = \text{héritabilité} ; r = \text{répétabilité} ; \lambda = \frac{4}{h^2} - 1 ; \gamma = \frac{2}{h^2} - 1$$

Information utilisée	Φ_{ij}	Expression de l'index
Performance individuelle P_i	1/2	$h^2(P_i - \mu)$
Moyenne de n performances individuelles \bar{P}_i	1/2	$\frac{nh^2}{1 + (n-1)r}(\bar{P}_i - \mu)$
Performance d'un parent direct P_j	1/4	$\frac{1}{2}h^2(P_j - \mu)$
Performance d'un grand-parent P_j	1/8	$\frac{1}{4}h^2(P_j - \mu)$
Moyenne \bar{P} de n descendants directs tous demi frères-sœurs entre eux (par ex. issus du même père et de mères différentes)	1/4	$2\frac{n}{n+\lambda}(\bar{P} - \mu)$
Moyenne de n demi frères-sœurs \bar{P}	1/8	$\frac{n}{n+\lambda}(\bar{P} - \mu)$
Moyenne de n pleins frères-sœurs \bar{P}	1/4	$\frac{n}{n+\gamma}(\bar{P} - \mu)$

4. L'évaluation combinée

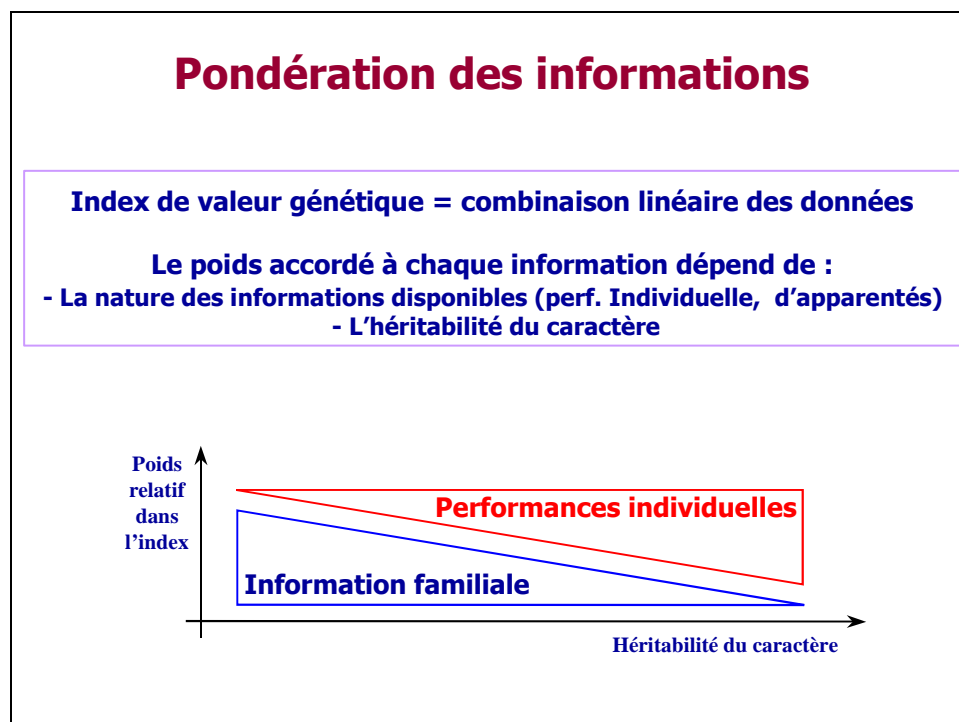
Dans la pratique, on dispose de plusieurs types d'information pour évaluer un même animal : performances des parents, individuelle(s), de collatéraux, etc. On réalise alors une évaluation combinée : au lieu de calculer autant d'index qu'il y a de types d'information différents (!), on calcule un index unique intégrant toute l'information disponible. Au lieu d'une régression linéaire simple, on effectue une régression linéaire multiple (voir Stat). Soit P_1, P_2, \dots, P_q les informations disponibles, l'index s'écrit :

$$\text{Index} = E(A_i | P_1, P_2, \dots, P_q) = b_1(P_1 - \mu) + b_2(P_2 - \mu) + \dots + b_q(P_q - \mu)$$

Dans cette approche, l'information utilisée (P) doit être synthétique. Comme vu plus haut, les performances d'apparentés du même type peuvent être regroupées dans une moyenne. Si les parents sont connus, on peut synthétiser l'information les concernant sous forme d'index de valeur génétique et calculer un *index sur ascendance*. Soit \hat{A}_p et \hat{A}_m les index du père (p) et de la mère (m), l'expression de l'index sur ascendance découle de la notion même d'additivité :

$$\text{Index sur ascendance} = E(A_i | \hat{A}_p, \hat{A}_m) = \frac{1}{2} \hat{A}_p + \frac{1}{2} \hat{A}_m$$

Les poids (b) accordés aux informations sont des coefficients de régression multiple. La détermination de ces poids fait intervenir (1) la liaison statistique entre chaque variable prédictrice (P) et la valeur à prédire (A) et (2) les liaisons statistiques entre les différentes variables prédictrices : si deux variables sont corrélées, statistiquement cela signifie que l'une est partiellement contenue dans l'autre et il convient donc de tenir compte de cette redondance. Toutes choses étant égales par ailleurs, moins le caractère considéré est héritable, plus le poids relatif accordé à l'information familiale (parents et collatéraux) par rapport aux performances individuelles est élevé. A l'inverse, quand le caractère est fortement héritable, les performances individuelles sont déjà suffisamment informatives et l'on accorde moins de poids relatif à l'information familiale.



C. Le mode d'expression des index de valeur génétique

Le mode d'expression actuel des index de valeur génétique est issu des usages qui se sont instaurés au cours du temps, de façon variable d'une espèce et d'une production et l'autre, et d'un pays à l'autre. Il est essentiel de connaître ces différences, notamment compte tenu de la dimension internationale que prend la sélection animale dans un nombre croissant d'espèces.

1. Valeur génétique ou différence attendue ?

Les résultats d'évaluation fournissent les valeurs génétiques estimées de l'ensemble des candidats à la sélection alors que c'est la valeur de la descendance future qui nous préoccupe. Sur la base de ce constat, les praticiens de plusieurs pays ont opté pour une expression des résultats non pas en terme de valeur génétique mais en terme de différence attendue chez les descendants, ce qui revient à effectuer une division par deux. En ce qui concerne l'espèce bovine, les pays d'élevage laitier développé en dehors de l'Europe continentale (USA, Canada, Israël, Royaume-Uni) expriment leurs résultats en différence attendue, alors que les pays d'Europe continentale publient la valeur génétique des reproducteurs eux-mêmes.

2. Choix d'un repère et éventuelles transformations

Les index de valeur génétique, tout comme les performances, s'expriment en unités du caractère considéré : nombre de porcelets sevrés par mise-bas, g de protéine par kg de lait, etc. La forme sous laquelle ils sont exprimés doit rendre facile, sinon attrayante, leur utilisation comme critère de sélection. Cela est particulièrement vrai pour les espèces où la sélection s'effectue en partie en ferme. En effet, dans cette situation, il convient de faciliter le choix des éleveurs auprès de qui les résultats d'évaluation génétique sont diffusés. A un instant donné, dans une population donnée, la valeur génétique est une variable centrée (d'espérance nulle). Il en découle que les index, calculés tel que défini plus haut, sont eux aussi centrés : ils s'expriment sous forme d'écart à une moyenne nulle et seuls sont qualifiés d'améliérateurs les animaux ayant un index positif. Pour des raisons de « lisibilité », on peut être amené à effectuer certaines transformations des index après calcul, transformations qui, évidemment, ne doivent pas changer le classement des candidats à la sélection.

a. Choix d'une base de référence

Lorsque l'évaluation génétique concerne simultanément un grand nombre d'animaux nés à des générations différentes, il est d'usage courant d'exprimer les index de valeur génétique non pas en écart à la moyenne de l'ensemble des animaux évalués mais en écart à la moyenne d'un sous-groupe aux caractéristiques bien précises, comme, par exemple, les animaux nés à une période donnée. Il ne s'agit là que d'un simple changement de repère.

Il est possible d'employer une *base fixe* dans le temps. Dans ce cas, les index sont toujours exprimés en écart à la moyenne du même groupe d'animaux de référence. Au sein des populations avec progrès génétique, les index des jeunes générations sont alors en moyenne supérieurs à ceux des générations plus anciennes. Afin de faciliter le tri instantané des reproducteurs, on peut utiliser une *base mobile*, c'est-à-dire exprimer les index en écart à une référence qui change d'une année sur l'autre. A titre d'exemple, c'est le choix qui a été fait en France pour l'édition des index chez les bovins, depuis la fin des années 1970 pour les races laitières (encadré ci-dessous) et depuis le début des années 1990 pour les races allaitantes. La comparaison d'animaux évalués à des dates différentes, c'est-à-dire dont les index sont exprimés en écart à des bases différentes, nécessite alors de connaître les valeurs successives des bases mobiles et de ramener tous les indices par rapport à la même base de comparaison.

b. Centrage et standardisation

L'évaluation génétique peut parfois porter sur des effectifs réduits, par exemple des séries de contrôle en station. Il est alors courant d'exprimer les index en écart à la moyenne de la série ou à la moyenne de quelques séries qui se sont succédées dans le temps. Cependant, pour faciliter le choix des éleveurs on cherche à ce que l'échelle des valeurs soit semblable d'une série à l'autre, par simples centrage et standardisation des index. Souvent, car il s'agit d'une échelle familière, on se ramène à une moyenne de 100 et on fait en sorte qu'un écart de 20 points corresponde à un écart-type (σ), phénotypique ou génétique, du caractère :

$$\text{Index diffusé} = 100 + 20 \frac{\text{Index calculé} - \text{Moyenne des index calculés}}{\sigma}$$

Définition des bases mobiles d'édition des index des bovins laitiers en France

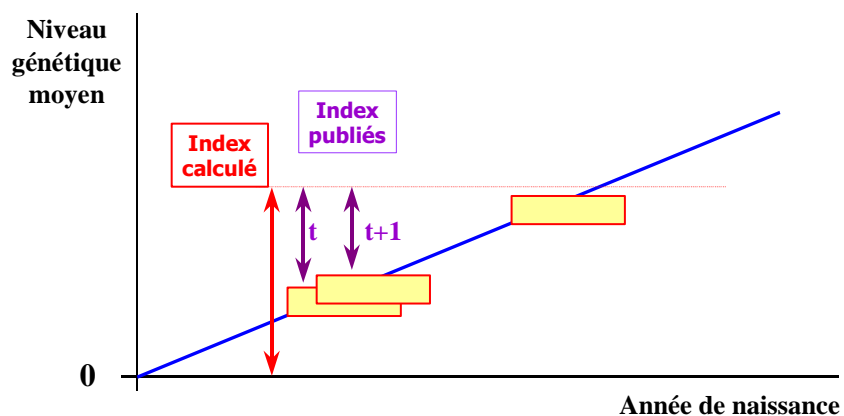
Base de calcul (= base fixe) : ensemble des vaches nées de 1976 à 1980 de parents connus et ayant fait l'objet d'un contrôle de performances. C'est par rapport à la moyenne de ce groupe de référence que sont calculés les index de tous les animaux, mâles et femelles confondus.

Base d'édition (= base mobile) est différente pour les mâles et les femelles. Pour les 3 principales races, les bases mobiles sont définies comme suit :

- mâles : base mobile de l'année t = moyenne (en base fixe) des index des taureaux d'IA nés durant les années $t-10$ à $t-7$.
- femelles : base mobile de l'année t = moyenne (en base fixe) des index des vaches contrôlées, nées l'année $t-6$.

Pour les autres races, du fait d'effectifs plus réduits, la valeur des bases mobiles est établie à partir de groupes définis sur des périodes de temps plus larges.

Illustrations des notions de bases de référence



D. Le Coefficient de Détermination

1. Définition

L'index représente la valeur génétique la plus probable compte tenu de l'information dont on dispose. Ce disant, on admet que l'on peut se tromper (mais la méthode employée garantit que l'on minimise l'erreur commise). En statistiques, on caractérise l'ampleur de l'incertitude par la variance d'erreur¹⁶ et, *a contrario*, le degré de confiance à accorder à la prédiction par le Coefficient de Détermination (*CD*)¹⁷.

Dans différents documents d'élevage (catalogue de reproducteurs d'insémination artificielle, fiche-carrière d'une femelle reproductrice, etc.), les index de valeur génétique doivent être accompagnés de leur *CD*¹⁸. Par définition, le *CD* est égal au carré du coefficient de corrélation entre l'index et la valeur génétique vraie¹⁹ :

$$CD = \left[\text{cov}(\hat{A}_i, A_i) / \sigma_{\hat{A}} \sigma_A \right]^2$$

La variance d'erreur attachée à la prédiction, c'est-à-dire l'ampleur des variations possibles de la valeur génétique vraie autour de l'index, s'exprime en fonction du *CD* :

$$\text{variance d'erreur de la prédiction} = \text{var}(A_i | \hat{A}_i) = (1 - CD) \sigma_A^2$$

Cette expression souligne que l'augmentation du *CD* représente une réduction d'incertitude sur la valeur génétique estimée.

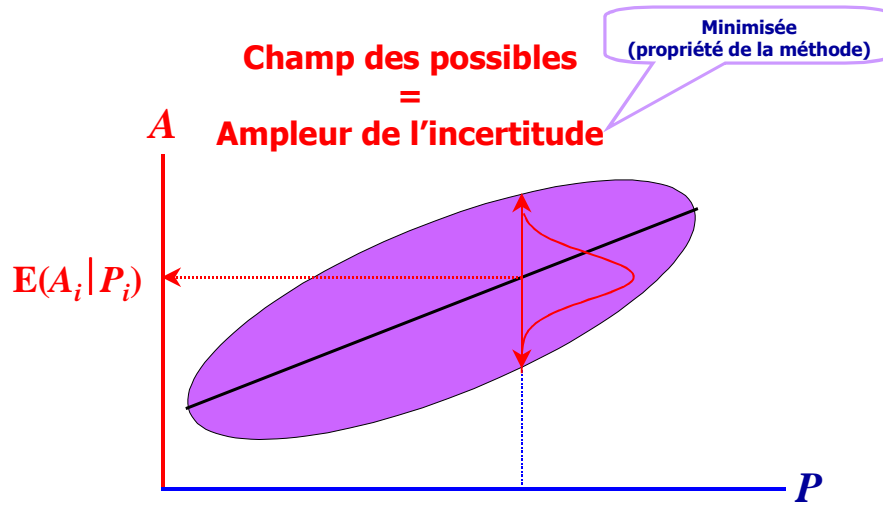
¹⁶ On évoque ici les erreurs aléatoires (c'est-à-dire non orientées) de prédiction et non pas en liaison avec d'éventuels biais (orientés dans un sens donné).

¹⁷ On dit également que le *CD* quantifie la *précision* de l'évaluation, terme que nous réserverons cependant pour décrire les paramètres du progrès génétique sous sélection (chapitre V).

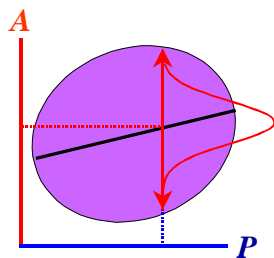
¹⁸ En anglais, le Coefficient de Détermination est désigné sous le terme de *repeatability*, abrégé par *Rep* et à ne pas confondre avec la répétabilité d'un caractère.

¹⁹ De ce fait, on emploie également "*R*²", mais le terme "*CD*" est devenu d'usage plus courant.

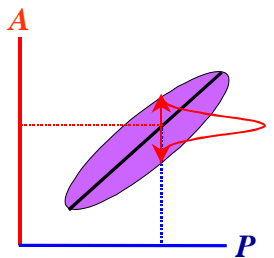
Prédiction n'est pas certitude



Ampleur de l'incertitude (variance d'erreur)



Forte incertitude
Faible Coefficient de Détermination



Faible incertitude
Fort Coefficient de Détermination

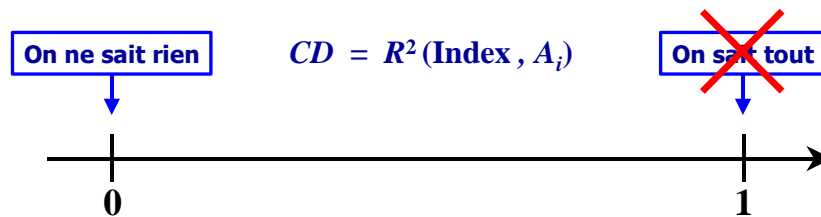
Coefficient de Détermination

- CD -

Degré de confiance à accorder à un index de valeur génétique

Une augmentation du CD traduit une réduction d'incertitude

Carré du coefficient de corrélation entre l'index et la valeur génétique vraie



Facteurs de variation du CD

- L'héritabilité (h^2) du caractère
- La nature et la quantité d'information prise en compte
 - Type de performance : individuelle ou d'apparenté
 - Nombre de performances
 - Corrélations entre les différentes performances

2. Facteurs de variation du *CD*

Dans la pratique, les valeurs génétiques vraies ne sont pas connues. C'est la formule de covariance entre apparentés qui fournit *a priori* les éléments du calcul du *CD* en fonction de la nature de l'information employée pour l'évaluation génétique et donc des liens de parenté entre les individus considérés. La valeur du *CD* dépend fondamentalement de deux facteurs :

- La valeur de l'héritabilité du caractère, le *CD*, toutes choses étant égales par ailleurs, croissant avec celle-ci.
- La nature et la « quantité » d'information prise en compte : certains apparentés sont plus informatifs que d'autres, le *CD* croît, mais pas de façon linéaire, avec le nombre d'apparentés considérés d'un type donné, certaines informations sont redondantes si bien que les *CD* partiels ne s'ajoutent pas, etc.

Le tableau 3 fournit quelques expressions « classiques » des *CD*.

- Le *CD* d'une évaluation sur une seule performance individuelle est égal à l'héritabilité du caractère. La connaissance de plusieurs performances individuelles induit un accroissement du *CD*, mais d'autant plus faible que le caractère est répétable (que la seconde performance ressemble à la première, etc.).
- A l'exception de l'évaluation sur descendance, tous les *CD* ont une valeur limite inférieure à 1 quand la « quantité » d'information (n) tend vers l'infini : h^2/r pour l'évaluation sur performances individuelles, $1/2$ pour l'évaluation sur ascendance (aléa de la méiose), $1/4$ ou $1/2$ pour l'évaluation sur collatéraux.
- L'évaluation sur descendance constitue l'épreuve « de vérité » : on peut obtenir le *CD* que l'on veut (tendant vers 1). Dans la pratique, on cherche un compromis entre une valeur élevée de *CD* et les coûts d'organisation : par exemple, pour les taureaux laitiers d'insémination, une valeur de *CD* de 0,7, obtenue avec la mesure des premières lactations d'environ 70 filles.

Tableau 3. Expression du *CD* des index de valeur génétique selon l'information considérée.

$$h^2 = \text{héritabilité} ; r = \text{répétabilité} ; \lambda = \frac{4}{h^2} - 1 ; \gamma = \frac{2}{h^2} - 1$$

Information utilisée	Expression du <i>CD</i>
Performance individuelle	h^2
Moyenne de n performances individuelles	$\frac{nh^2}{1 + (n-1)r}$
Performance d'un parent direct	$\frac{1}{4}h^2$
Performance d'un grand-parent	$\frac{1}{16}h^2$
Index des parents (\hat{A}_p et \hat{A}_m)	$\frac{1}{4}CD_p + \frac{1}{4}CD_m$
Moyenne de n descendants demi frères-sœurs entre eux	$n/(n + \lambda)$
Moyenne de n demi frères-sœurs	$\frac{1}{4}[n/(n + \lambda)]$
Moyenne de n pleins frères-sœurs	$\frac{1}{2}[n/(n + \gamma)]$

E. L'évolution de la valeur d'index et du *CD* dans la vie d'un animal

L'évaluation génétique est effectuée à intervalles réguliers, selon un rythme mensuel (porc) à annuel (chevaux). A chaque évaluation, on conserve toute l'information dont on disposait auparavant et de nouvelles informations sont apportées : nouveaux animaux en production, nouvelles performances d'animaux déjà présents, nouveaux descendants, etc. Par ailleurs, un animal fait partie d'une population dont la moyenne génétique évolue graduellement. Une conséquence fondamentale d'une telle situation et d'un dispositif périodique d'évaluation est que l'index de valeur génétique d'un animal donné est susceptible d'évoluer dans le temps.

1. Evolution de la valeur d'index liée à l'évolution de la population

L'évolution du niveau génétique moyen de la population peut se traduire par le changement de la base de référence par rapport à laquelle les index sont exprimés (cf § C.2.a) : dans le cas d'une base mobile, les index individuels peuvent ainsi être modifiés d'une édition à l'autre.

Ce phénomène touche tous les animaux à la fois. Par exemple, en cas de progrès génétique (c'est ce que l'on cherche) et avec un chagement de base annuel, les animaux constituant le groupe de référence pour l'année $t+1$ étant en moyenne plus jeunes que ceux du groupe de référence pour l'année t , leur valeur génétique moyenne est meilleure. Le passage de l'année t à l'année $t+1$ se traduit donc par une diminution des index d'une quantité égale à la différence entre les valeurs des deux bases mobiles d'édition. Cela a une conséquence importante dans les populations où les reproducteurs mâles sont majoritairement utilisés par insémination artificielle (bovins et ovins laitiers, porcs) : sauf fluctuation individuelle (voir prochain §), les index de ces reproducteurs baissent d'année en année, ce qui encourage les éleveurs à utiliser les reproducteurs les plus jeunes, ce qui est favorable à la diffusion du progrès génétique.

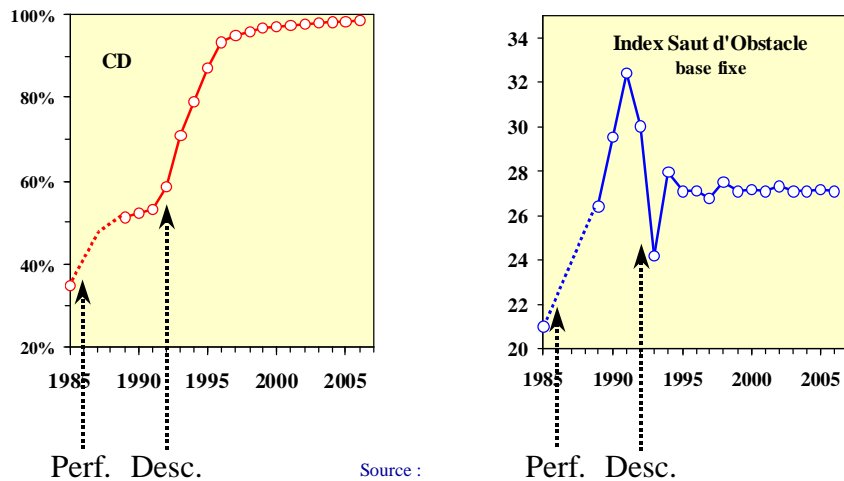
2. Evolution liée à des informations nouvelles de la valeur d'index d'un animal et du *CD*

Tout au long de sa carrière, l'information relative à un animal évolue : à sa naissance, on ne connaît que la valeur de ses parents, puis on connaît la (les) performance(s) qu'il réalise lui-même et, parfois, on peut connaître les performances de plusieurs de ses descendants alors qu'il est encore vivant. Dans le cadre d'une évaluation génétique en routine, on combine toutes ces informations. Cela a deux conséquences (voir diapositive suivante) : (i) selon que les nouvelles informations confirment ou non les informations précédentes, la valeur d'index peut fluctuer et (ii) au fur et à mesure que l'information s'enrichit, le *CD* augmente.

L'erreur de prédiction que l'on commet, et qui est inévitable, est plus importante au début qu'à la fin de la vie d'un animal : l'index calculé à la naissance (index sur ascendance) est moins corrélé à la valeur génétique vraie que ne l'est un index qui incorpore également les propres performances et, *a fortiori*, les performances de descendants. En d'autres termes, la valeur d'un animal en tant que reproducteur s'établit progressivement au fur et à mesure que l'on accumule de l'information à son sujet. Seule l'information sur un nombre minimal de descendants permet d'obtenir un *CD* très élevé. Cela ne veut en aucun cas dire que les autres informations sont inutiles : elles sont indispensables pour permettre un tri précoce parmi les candidats à la sélection ; l'évaluation génétique qui en résulte est simplement moins précise.

Evolution dans le temps de l'index d'un animal et du *CD* correspondant

Exemple d'un étalon Selle Français



Les malheurs de Quidam de Revel

Quidam de Revel est un étalon Selle Français né en 1982. Les performances en concours de saut d'obstacle sont centrées et standardisées : la moyenne est fixée à 100 et 20 points représentent un écart-type. Les indices génétiques, calculés et diffusés par l'INRA et les Haras Nationaux, sont exprimés par rapport à une base fixe constituée par la moyenne des animaux nés en 1974.

Quidam de Revel est un étalon « bien né » : son index d'ascendance était, en 1987, de + 21, avec un *CD* de 0,35, ce qui le situait dans les 5 % meilleurs de sa génération. Les résultats en compétition de Quidam de Revel sont élogieux : la moyenne de ses performances de 1986 à 1991 était de 151, soit 2 écarts-types et demi au dessus de la moyenne. L'index incorporant l'ascendance et les performances propres valait + 32 en 1991, avec un *CD* de 0,53. Une telle valeur, situant ce cheval parmi les tous meilleurs, justifiait pleinement qu'on le choisisse comme étalon reproducteur et qu'on l'utilise comme tel.

Hélas, les résultats en compétition des premiers descendants de Quidam de Revel n'ont pas confirmé les espoirs que l'on avait placés en lui : en 1993, la performance moyenne de ses 23 descendants se situait à 98, soit très légèrement en dessous de la moyenne générale. L'index calculé en 1993, qui combine l'ensemble des informations valait 24, soit 8 points de moins que l'index calculé en 1991, avec un *CD* de 0,71. Cette baisse, qui à l'époque avait fait couler de l'encre dans la presse spécialisée, s'explique en fait très bien par les résultats tout à fait moyens de la descendance. Si Quidam de Revel est un bon « performeur », il n'est sans doute pas si bon reproducteur que ses propres performances semblent l'indiquer.

En général, l'évolution, à la hausse comme à la baisse, des index calculés est liée à des phénomènes simples. Tout d'abord, du fait de l'aléa de la méiose, les valeurs génétiques des descendants d'un même couple de parents présentent une importante variabilité entre eux, ce qui peut expliquer les écarts constatés avec les index calculés sur ascendance. Ensuite, des performances individuelles peuvent s'écarter notablement de la moyenne essentiellement pour des raisons de milieu alors que l'on attribue à la valeur génétique une part (égale à h^2) de cet écart de performance, part qui est considérée comme un paramètre moyen. Enfin, du fait du hasard de l'échantillonnage, ou, plus grave, pour cause de traitements préférentiels non corrigés, le premier échantillon de descendants connus d'un reproducteur peut avoir des performances moyennes qui diffèrent significativement d'un échantillon plus grand.

