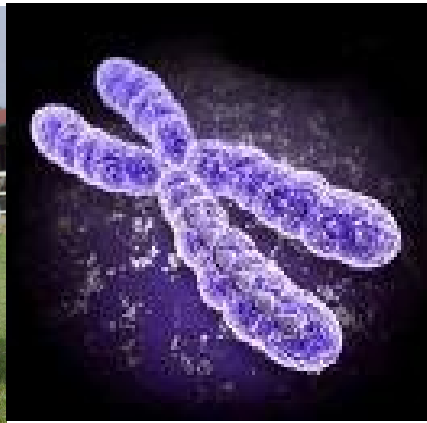


Principes des évaluations génétiques fondées sur des informations génomiques

F. Guillaume (I.E.), S. Fritz (UNCEIA), A. Baur (UNCEIA), S. Mattalia (I.E.)
P. Croiseau (INRA), JJ. Colleau (INRA), D. Boichard (INRA), V. Ducrocq (INRA)



INRA

UNCEIA

INSTITUT DE L'ÉLEVAGE

SAM²

Modèles sous-jacents aux indexations

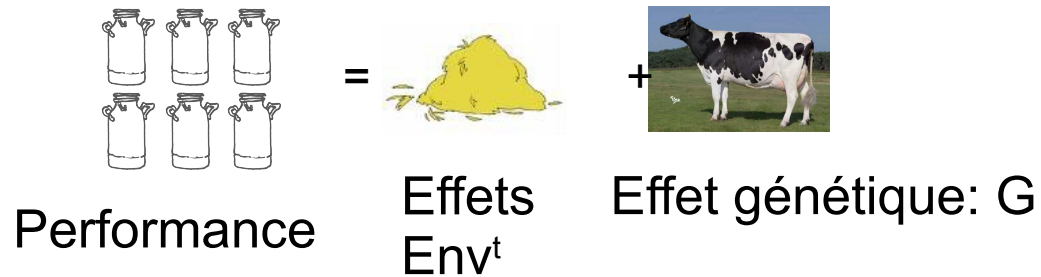
Utilisations de marqueurs pour l'évaluation

Méthodes d'évaluation génomiques

Validations des méthodes

Perspectives

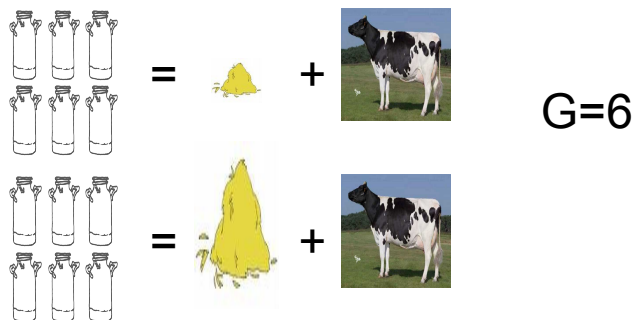
Modèle de description des performances



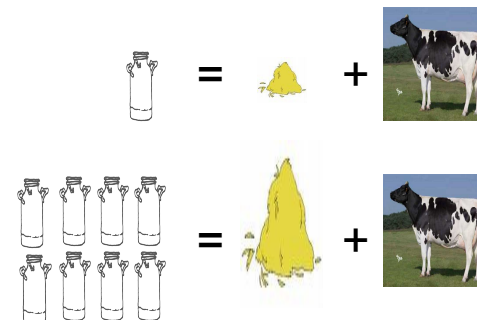
Si une performance est peu influencée par l'effet génétique, on parlera d'un caractère peu héritable

Si un caractère est peu héritable, l'estimation de l'effet génétique est plus difficile

Caractère héritable

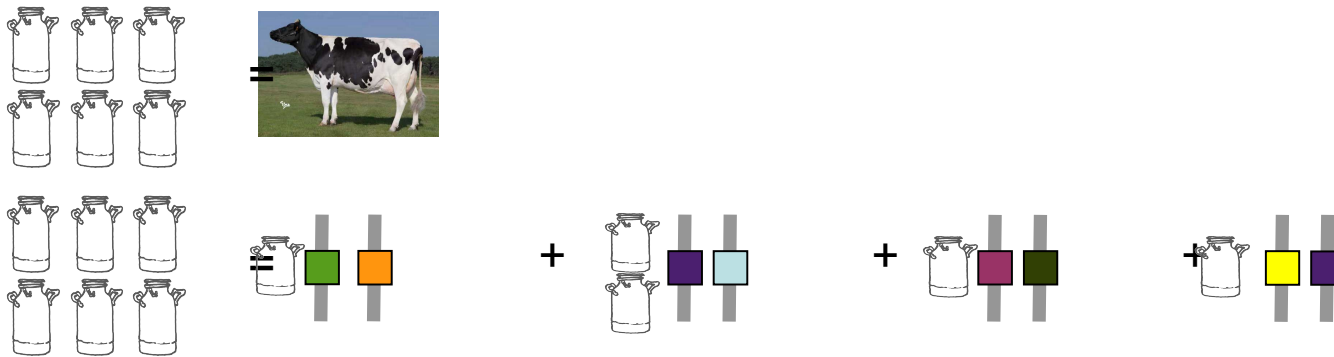


Caractère faiblement héritable



Valeur génétique additive

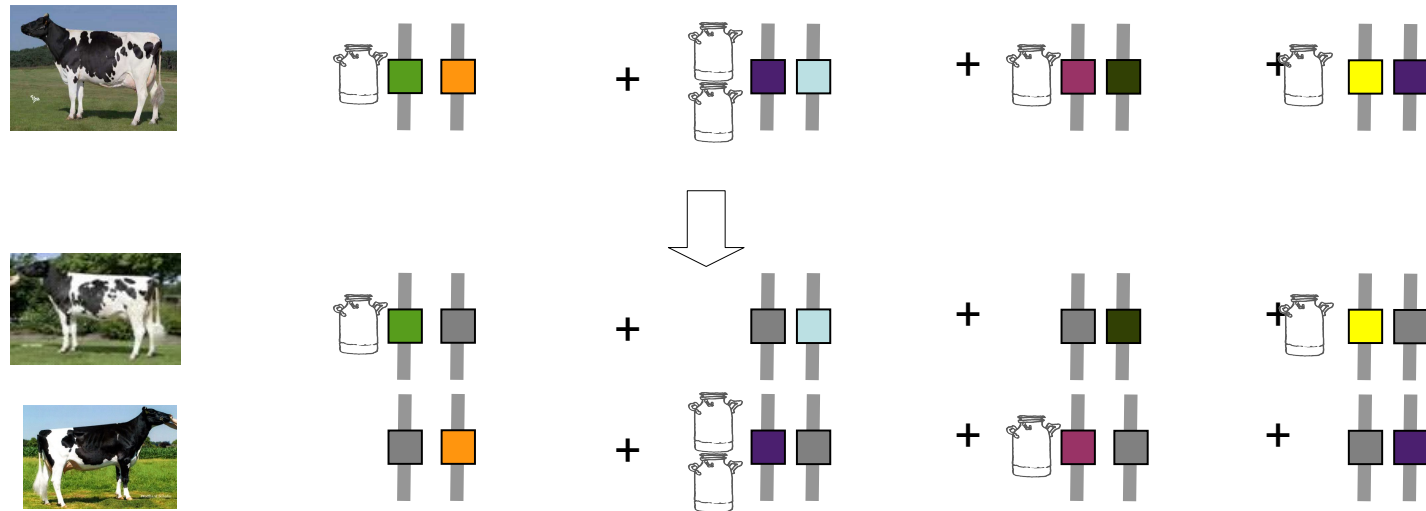
La valeur génétique est la somme d'effets de gènes portés par l'animal



Dans l'évaluation on pose l'animal comme étant l'un des facteurs explicatifs des performances. Or il s'agit bien des **gènes** qu'il porte qui expliquent les performances

Valeur génétique additive

Lors de la reproduction une **moitié** des gènes est **transmis aléatoirement** à la descendance



La moitié des gènes portés par un animal est également portée par sa mère. Il existera donc un ressemblance entre les performances d'un parent et celles de ses descendants

Modèle génétique polygénique

Selon le modèle polygénique, l'effet génétique est la somme des effets d'une infinité de gènes qui influencent un caractère d'intérêt

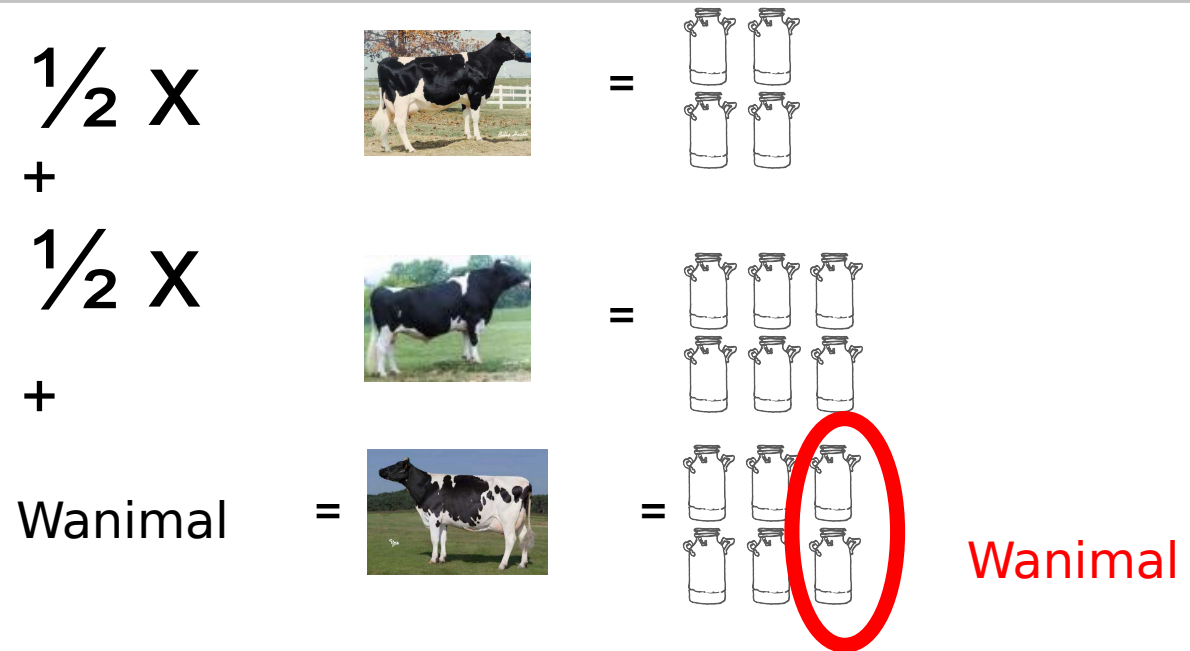
Ces gènes sont transmis pour moitié à la descendance

On pose l'équation suivante :

$$G_{\text{animal}} = \frac{1}{2} (G_{\text{père}} + G_{\text{mère}}) + W_{\text{animal}}$$

W_{animal} : est l'aléa de méiose : écart entre la moyenne des valeurs génétiques des parents et la valeur génétique vraie du descendant

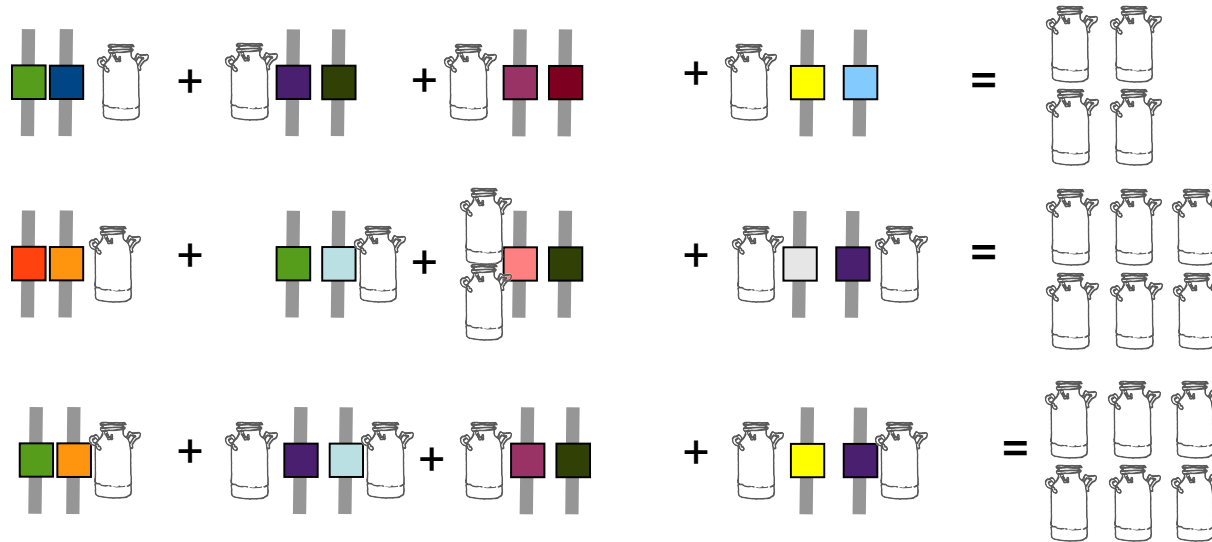
Évaluation génétique cas polygénique



En l'absence de performances,
 une part de la valeur génétique (w_{animal}) reste inconnue
 (et nulle en espérance)

On ne pourra estimer l'aléa de méiose
 qu'après avoir observé une performance

Principe de l'évaluation génomique



Si on avait une connaissance des gènes portés par les animaux auteurs des performances, on aurait un **accès direct** à leurs **valeurs génétiques vraies**

Avec l'approche de sélection génomique la ressemblance entre individus découle des gènes qu'ils auront en commun

Utilisation de marqueurs pour l'évaluation

Les gènes influençant les performances **sont rarement connus**

À défaut, on peut utiliser des **marqueurs proches** pour refléter les effets des gènes

Puces de 54 000 SNP

On espère identifier des associations entre allèles aux marqueurs et allèles aux gènes

FIGURE 1: BOVINESNP50 BEADCHIP

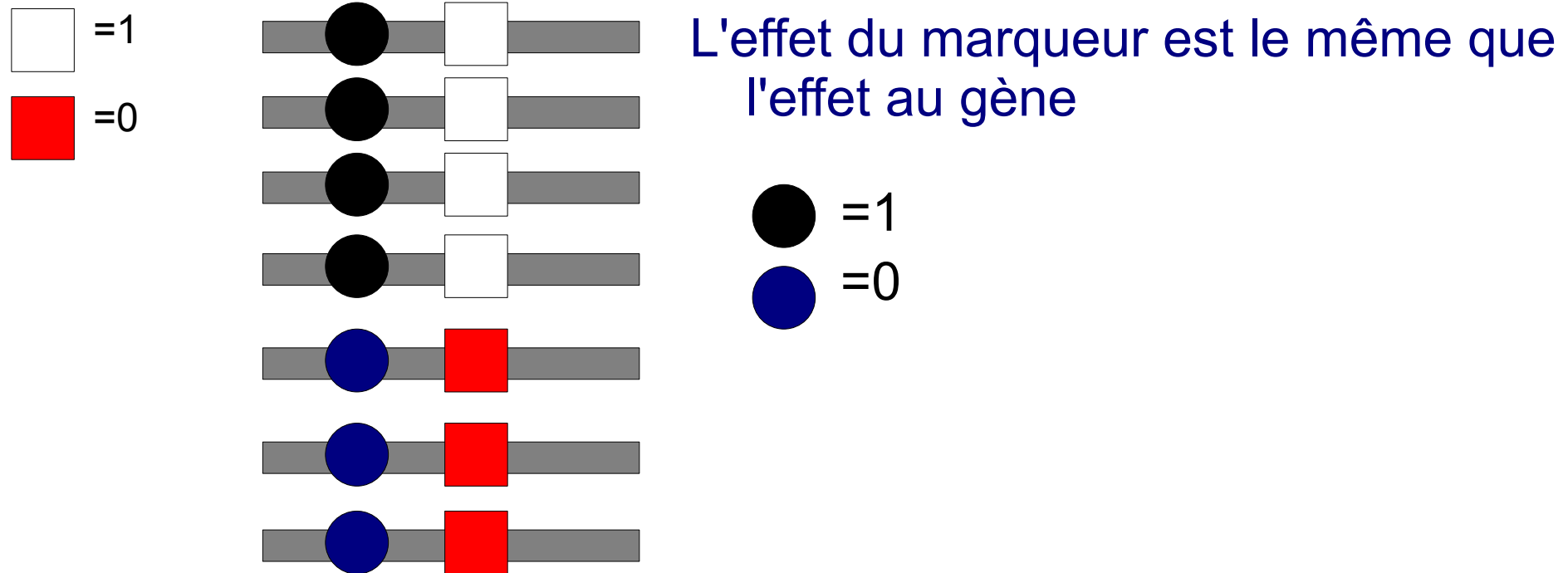


The BovineSNP50 BeadChip features more than 54,000 evenly-spaced SNPs across the entire bovine genome.

Utilisations des marqueurs : Illustrations

Cas idéal :

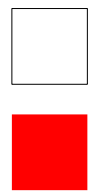
=> Un gène bi-allèlique toujours associé à un allèle au marqueur



Utilisations des marqueurs : Illustrations

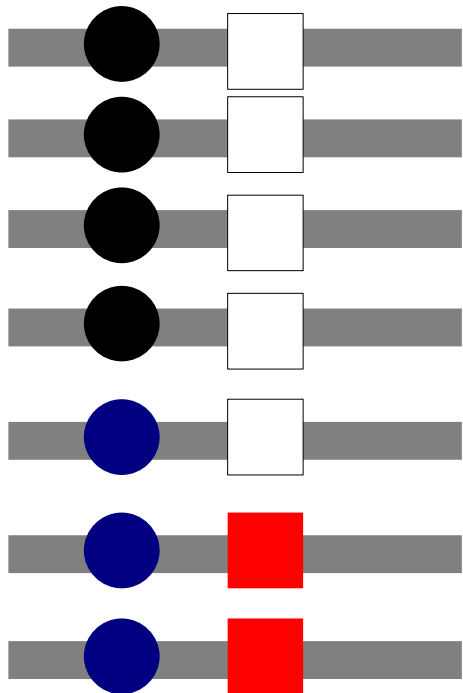
Cas plus réaliste...

Association incomplète

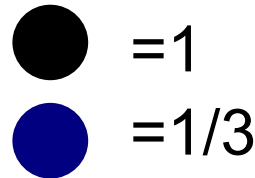


=1

=0



L'effet du marqueur ne traduit pas exactement l'effet du gène



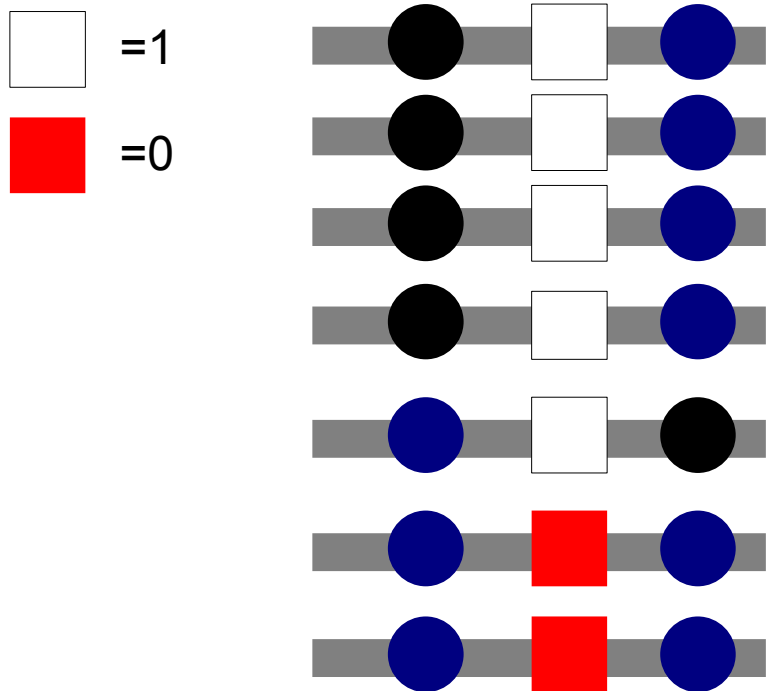
=1

=1/3

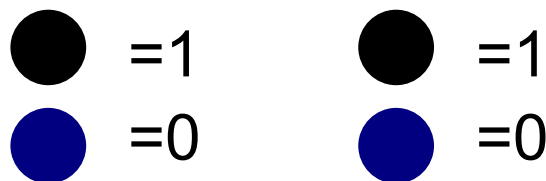
Utilisations des marqueurs : Illustrations

Cas plus réaliste...

Association incomplète



Les effets des marqueurs se compensent pour donner une estimation de l'effet des gènes



Déséquilibre de liaison

L'utilisation de marqueurs pour l'évaluation se base sur l'existence d'associations non aléatoires entre allèles aux marqueurs et allèles aux gènes : Déséquilibre de liaison

Ces associations existent du fait de :

La proximité entre marqueurs et gènes

Les effets de dérives et de sélection

Le déséquilibre de liaison peut s'établir de façon fortuite. On doit veiller à identifier les associations qui se maintiendront au fil des générations

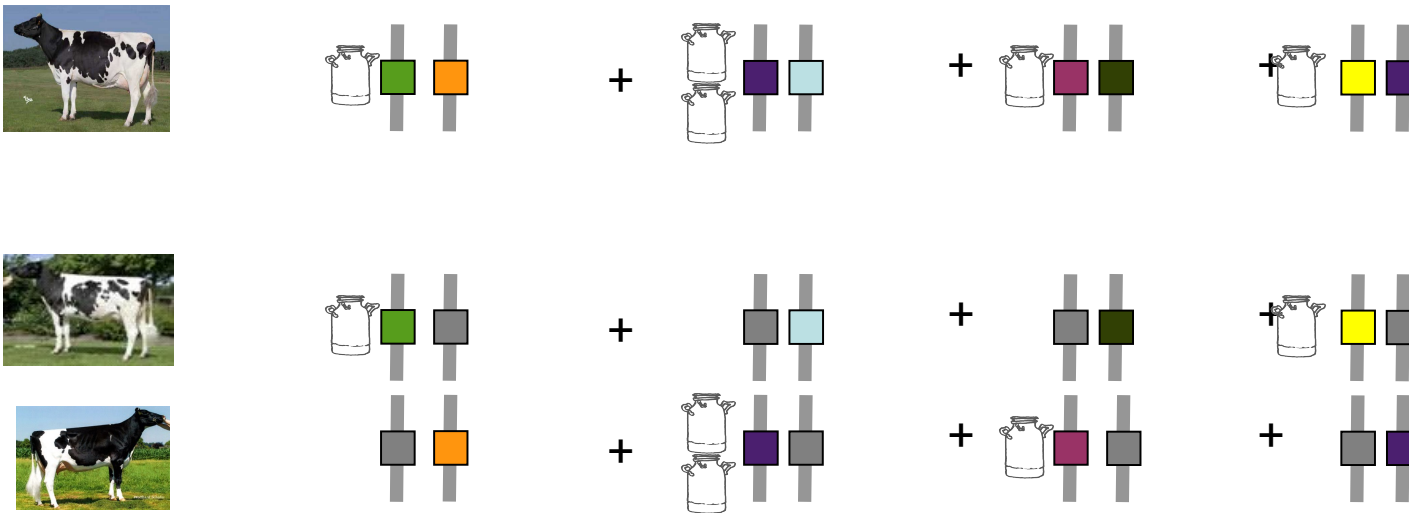
La puce 54K est un outil permettant a priori d'avoir un suivi efficace des régions contenant des gènes d'intérêts et d'identifier des associations entre allèles

La population observée doit d'être suffisamment importante et diversifiée pour ne pas identifier des associations manquant de généralité

Une autre difficulté réside dans l' adéquation entre le modèle (et ses paramètres), le caractère évalué (et son déterminisme génétique) et la structure de la population

Blup Génomique

On calcule la proportion réelle de gènes communs à deux individus pour estimer la ressemblance entre individus



Ici, ces deux sœurs n'ont aucun allèle aux gènes en commun ! La matrice de parenté génomique tiendra compte de cette situation

Blup Génomique

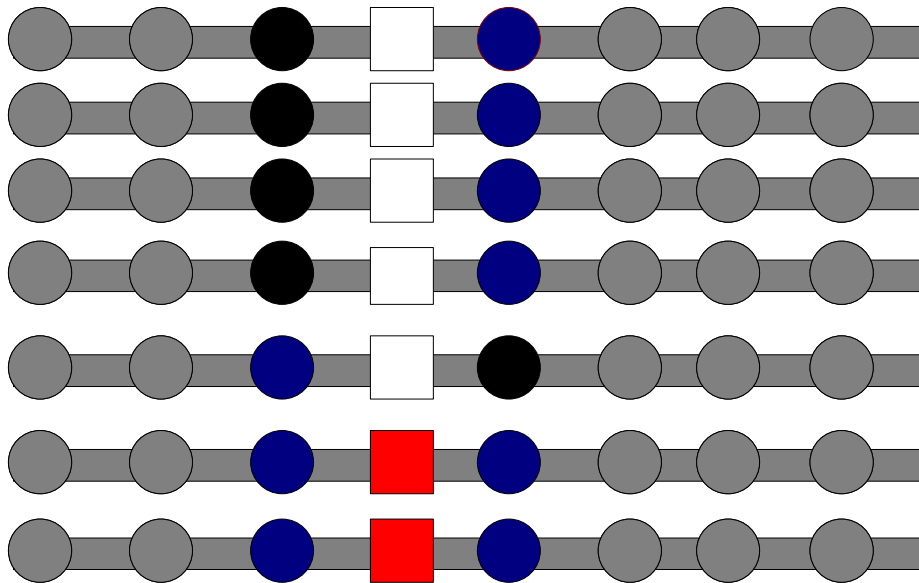
Méthode utilisée au USA, Canada, Allemagne

Cette méthode donne des calculs de la précision (CD) ayant tendance à être sur-estimés

Aucun a priori sur la localisation des marqueurs en début d'évaluation, on ne tient pas compte de la connaissance de gènes majeurs

Méthodes génomiques avec pré-sélection

On élimine les marqueurs sans effets pour ne conserver que les plus pertinents (Elastic-net)

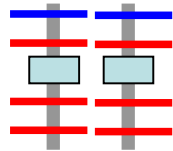


On estime les effets des marqueurs conservés

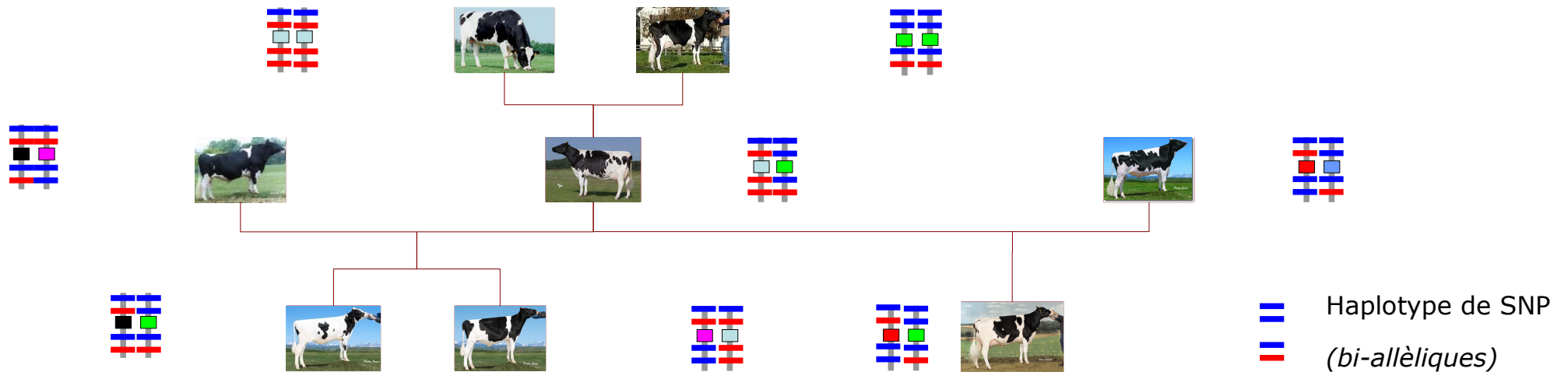
On somme les effets des allèles aux marqueurs retenus pour obtenir les valeurs génétiques individuelles

Détection de QTL (identification des régions chromosomiques ayant un effet important sur le caractère)

Sélection des marqueurs permettant le suivi des régions QTL en formant des haplotypes (séquence d'allèles présent sur un même chromosome à des marqueurs consécutifs)



À chaque évaluation on ré-estime les effets des haplotypes conjointement avec un effet polygénique



Les régions QTL sont suivies étroitement à l'aide d'haplotypes de SNP (~4-8)

Les effets des haplotypes traduisent l'effet l'allèle d'un gène sous jacent

On conserve un effet polygénique (expliquant une infinité de gènes chacun de faible effet)

Méthodes de validation (1)

Simulations

La valeur simulée est comparée à valeur génétique vraie

Résultats dépendants des hypothèses de simulations et des hypothèses du modèle

Il semble que le paramétrage correct des modèles soit assez difficile

Populations simulées souvent simplistes

Méthodes de validation (2)

Évaluation rétrospective

On se met en situation réaliste

Même qualité d'information phénotypique

Même structure de la population

Utilisation de l'info marqueur SNP

Correspond à la qualité de l'évaluation en début de programme ! (Info encore limitée)

Juge de paix pas fiable à 100 %

Méthodes de validation (3)

Re-sampling

On divise la population en 2 (e.g 90%/10%)

Utilisation de 90 % des infos pour évaluation équation de prédiction

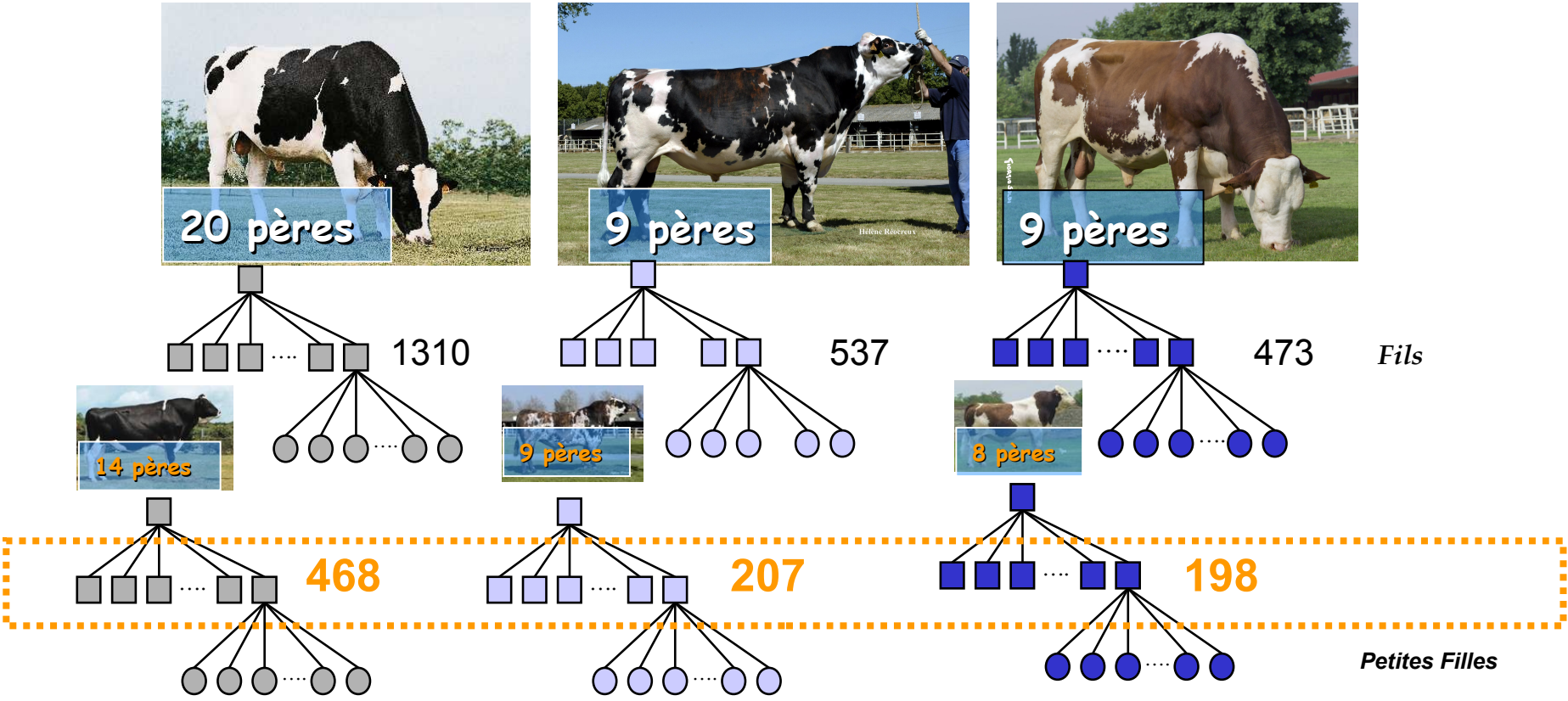
Vérification sur les 10 % de la qualité prédictive des équations

Itération du processus

La structure de population ne reflète pas celle des candidats

(qualité info phénotypique irréaliste)

Population de validation



Animaux sans performances en 2004, évalués après testage sur descendance en 2008, génotypés sur la puce 54 k

Validation en situation de 2004 (simulation + données réelles)

Illustration Lait Holstein

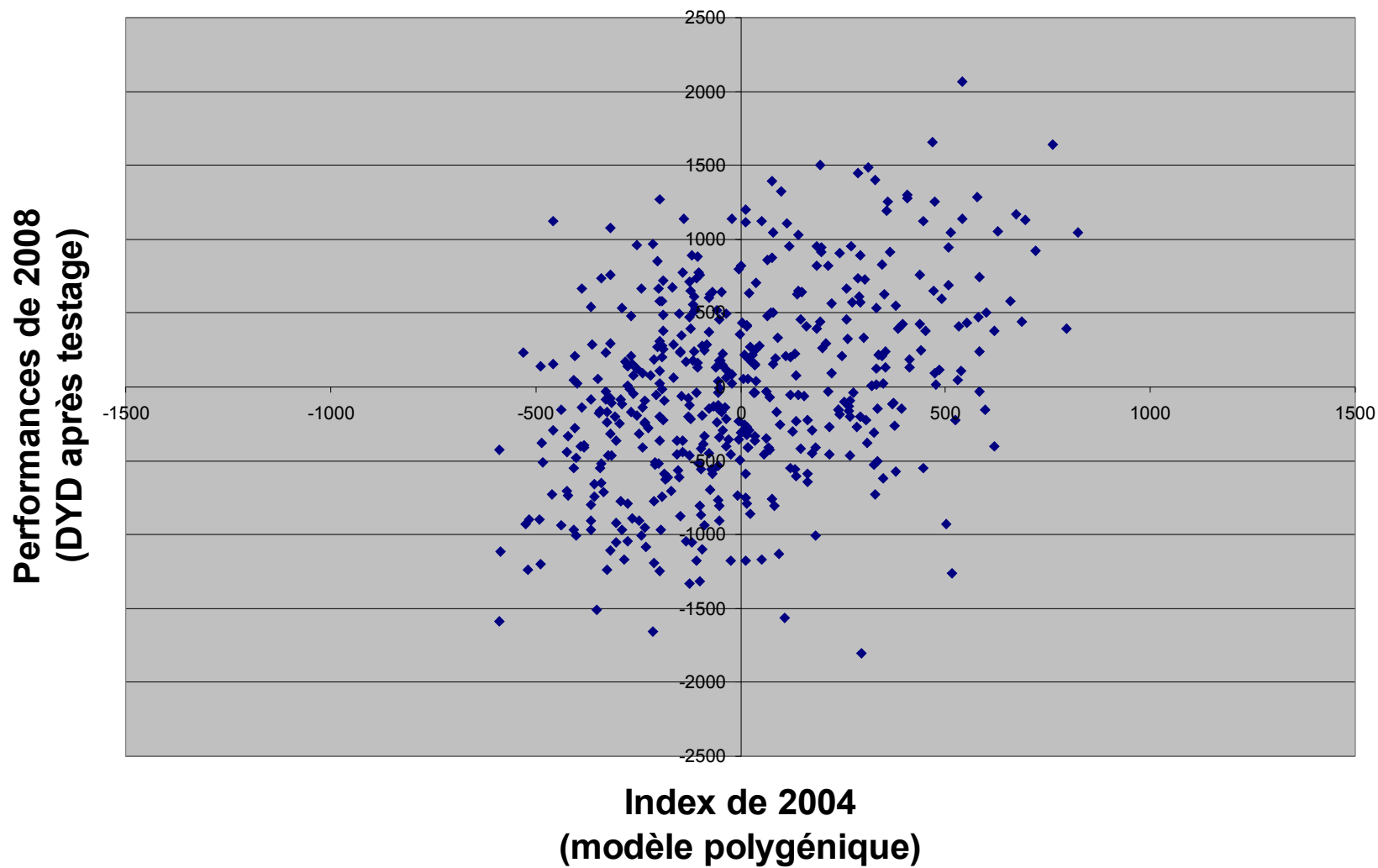


Illustration Lait Holstein

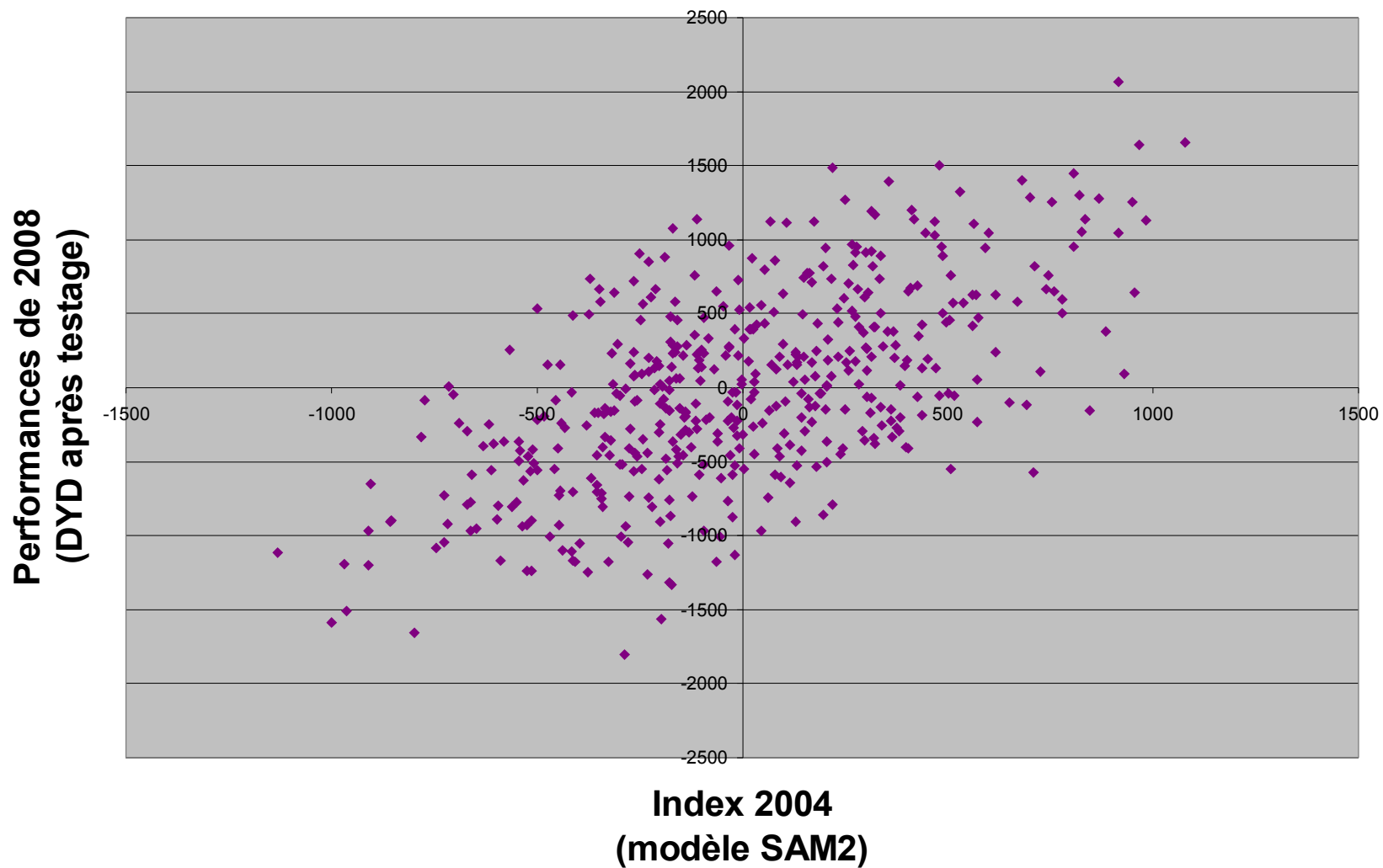
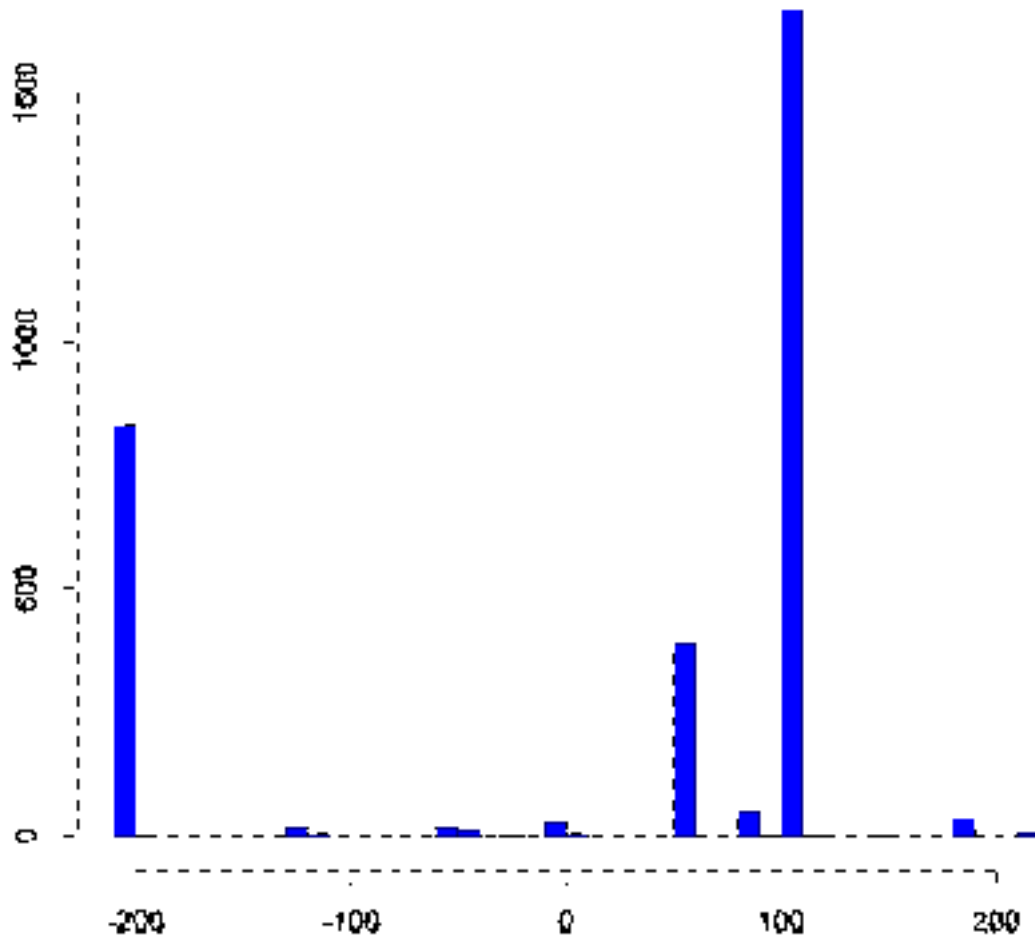


Illustration sur données réelles (cas de DGAT1)

Distribution des effets haplotypiques autour de DGAT1



Haplotypes de 4 SNP

Très peu d'haplotypes dans la population.

2 haplotypes principaux dont les effets semblent correspondre à ceux des allèles connues de DGAT1

La méthode SAM2 permet d'obtenir une meilleure prédiction des valeurs génétiques après testage sur descendance

Pour les gènes connus, les haplotypes utilisés permettent de capturer l'effet des gènes

La méthode présente de bons résultats même pour la Montbéliarde et la Normande

Les gains sont importants même pour les caractères à faible héritabilité

Approches Méthodologiques et Applications de la Sélection GENomique) projet financé par l'ANR, ayant pour but de :

Comparer les résultats de SAM2 à différentes méthodes de sélection génomique

Faire évoluer l'évaluation génomique

Étudier les biais potentiels dus aux évaluations génomiques

Comparaisons de différentes méthodes de sélection génomiques

Corrélations entre performances après testage sur descendance et index avant testage obtenus avec différentes méthodes d'évaluation pour des candidats Montbéliards

Montbéliarde	Modèle Polygénique	SAM2	Elastic Net (avec carto)	Elastic Net	GBLUP (avec carto)	GBLUP
Lait	0,273	0,420	0,493	0,362	0.35	0.49
Matière Protéique	0,276	0,383	0,549	0,374	0.41	0.46
Matière Gras	0,355	0,438	0,469	0,404	0.46	0.46
Taux Protéique	0,214	0,543	0,392	0,508	0.41	0.48
Taux Butyreux	0,372	0,579	0,614	0,364	0.34	0.51

Premiers enseignements de AMASGEN

La pré-sélection de SNP présents dans des régions QTL permet d'obtenir les évaluations les plus précises

Les résultats se situent dans le même ordre de valeurs

La méthode donnant les meilleurs résultats n'est pas la même selon les caractères étudiés

=> Comment évolueront ces résultats avec une base de données plus importante ?

Projet Eurogenomics

Partage des populations de référence en race Holstein:

FRA, ALL, NLD, DNK

Objectifs :

Atteindre une taille critique plus rapidement

Avoir des évaluations génomiques les plus précises possibles

Aller plus loin dans les collaborations scientifiques au niveau européen

La concurrence entre ES ne se situe pas dans les outils d'évaluations !

Évaluation dans les races à petits effectifs

La quantité de phénotypes disponibles est plus faible que dans les grandes races

Besoin d'une approche alternative:

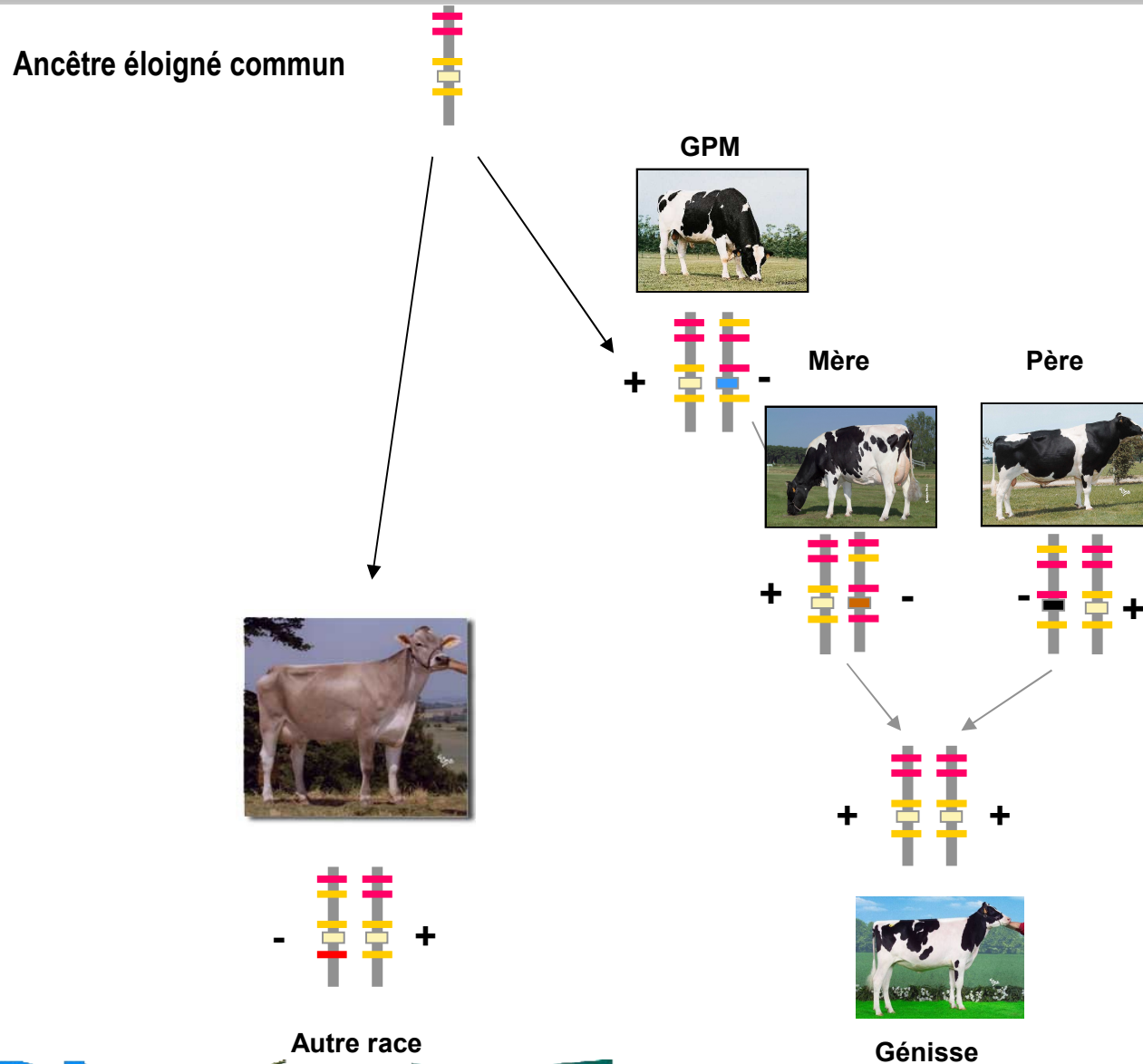
Puce 600 K SNP

Identification d'association entre allèles aux marqueurs et allèles aux gènes vérifiée dans toutes les races

Les phénotypes d'animaux des différentes races permettent une localisation plus précise des QTL

A partir des résultats de détection de QTL multi-races (et vérification de l'existence des QTL) , une évaluation assistée par marqueurs est envisageable

Évaluation dans les races à petits effectifs



Évaluation dans les races à petits effectifs

Implique de commencer à collecter au plus vite de l'ADN d'animaux ayant des performances !

En routine, une puce très dense (600 K) n'est pas forcément nécessaire

D'autres pistes : se regrouper entre pays (Projet Interbull Brun)

Les nouveaux outils de génotypages permettent de réinterpréter le modèle polygénique pour évaluer la valeur génétique additive

Pour des animaux sans performances, ces modèles permettent des évaluations plus précises

Les gains présentés reposent encore sur des échantillons de données de faible taille (ne permettant pas forcément une estimation précise de l'ensemble des effets et paramètres du modèle). La collecte continue de données phénotypiques et génotypiques est donc encore nécessaire

L'évaluation assistée par marqueurs n'est qu'un **outil**, les vrais enjeux de la sélection génomique sont dans l'usage de cet outil :

Progrès génétique dans les caractères jusqu'ici difficiles à sélectionner (Fonctionnels)

Meilleure gestion des populations de reproducteurs

Intégration de nouveaux caractères pour répondre aux nouvelles attentes

Merci pour votre attention

Principes des évaluations génétiques fondées sur des informations génomiques

F. Guillaume (I.E.), S. Fritz (UNCEIA), A. Baur (UNCEIA), S. Mattalia (I.E.)
P. Croiseau (INRA), JJ. Colleau (INRA), D. Boichard (INRA), V. Ducrocq (INRA)

