



La sélection génomique chez les bovins laitiers

Didier Boichard



12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

La situation dans les espèces d'élevage

- Après l'homme (2001) et la souris, séquençage des génomes complets de la poule (2004), du chien (2005), du bovin (2006), du cheval (2007), porc (2009), mouton, lapin...)
- Entièrement dans le domaine public
- Séquençage comparé de différents individus => mise en évidence des différences (polymorphismes)
3,2 millions de polymorphismes bovins dans dbSNP
- Nouvelles technologies de génotypage par puce

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

La puce SNP Illumina bovine

- 54 000 SNP répartis de façon homogène sur l'ensemble du génome bovin
- A un coût raisonnable

FIGURE 1: BOVINESNP50 BEADCHIP



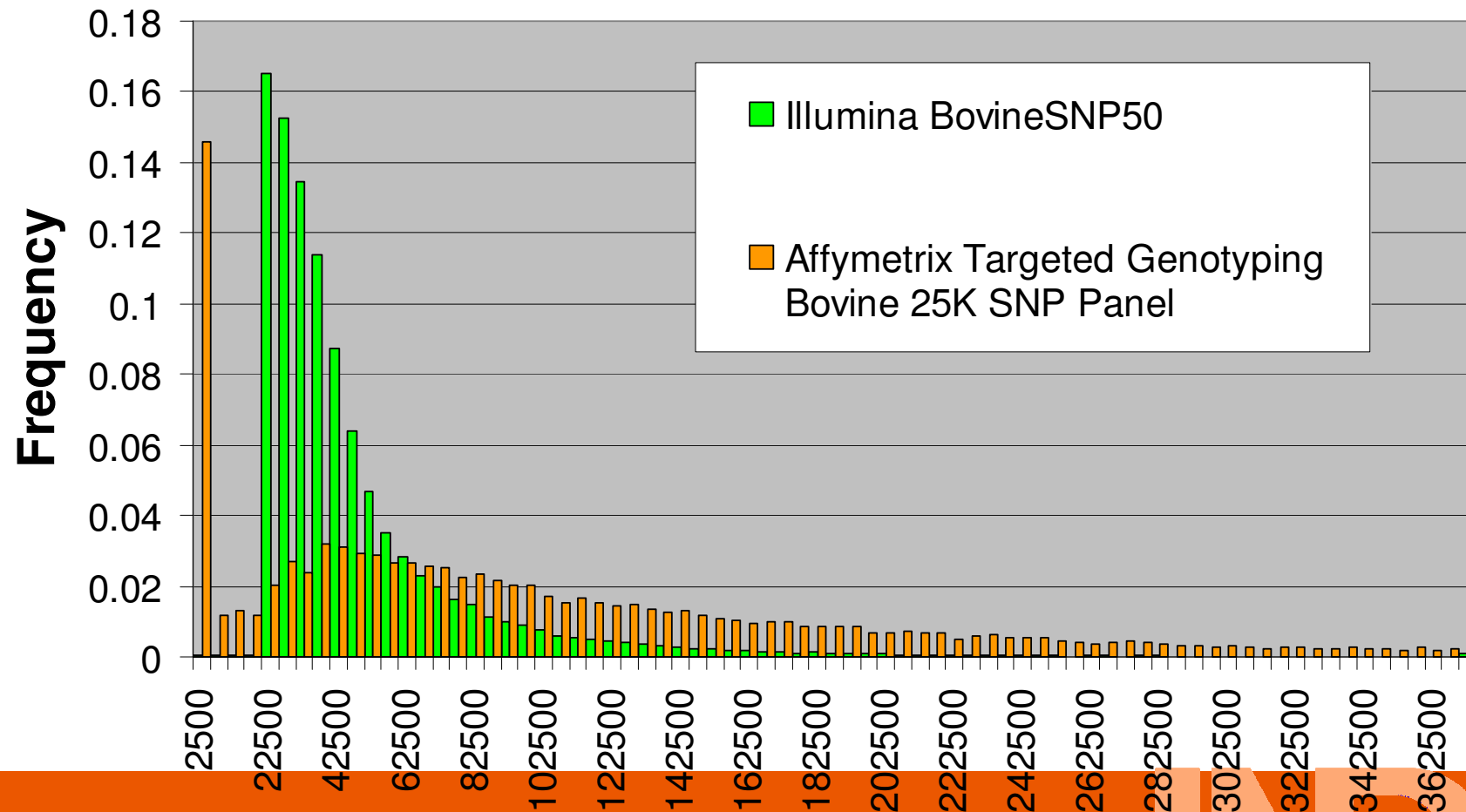
The BovineSNP50 BeadChip features more than 54,000 evenly-spaced SNPs across the entire bovine genome.

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Distribution des intervalles entre SNP



Illumina Bovine SNP50
Mean Interval : 45,878 Kb

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA



12 mai 2009

TABLE 3: BOVINESNP50 BEADCHIP CONTENT VALIDATION

BREED	SAMPLES	POLYMORPHIC LOCI*	MEAN MAF†	MEDIAN MAF†
Angus	60	41,491	0.21	0.21
Beefmaster	24	42,925	0.22	0.21
Bos indicus Gir	24	23,971	0.11	0.02
Bos indicus Nelore	21	25,814	0.11	0.02
Brahman	25	30,284	0.13	0.08
Brown Swiss	24	36,347	0.19	0.17
Charolais	26	42,589	0.22	0.21
Guernsey	21	38,632	0.19	0.17
Hereford	32	42,992	0.20	0.23
Holstein	64	42,730	0.22	0.22
Jersey	28	35,976	0.18	0.14
Limousin	45	42,821	0.22	0.22
N'Dama	25	29,049	0.14	0.08
Norwegian Red	21	42,782	0.22	0.21
Piedmontese	24	42,185	0.22	0.21
Red Angus	15	40,188	0.21	0.20
Romagnola	24	38,830	0.20	0.19
Santa Gertrudis	24	42,064	0.22	0.21
Sheko	20	35,726	0.17	0.12
Outgroup‡	18	11,206	0.05	0.00
Overall	565	47,545	0.25	0.24

*MAF > 0.05

†Across all 54,001 loci

‡*Bos bison*, *Bos gaurus*, *Bos grunniens*, *Bos javanicus*, *Bubalus depressicornis*, and *Syncerus caffer*.

**Des SNP
suffisamment
polymorphes
pour être
informatifs
dans
différentes
races**



Définition

Sélection génomique :

sélection basée sur une estimation de valeur génétique des candidats à partir de l'information de marqueurs denses couvrant tout le génome

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



Comment ça marche ?

La méthode princeps : Meuwissen et al, 2001

Valeur génétique = somme des effets estimés de tous les marqueurs

Effets estimés avec une méthode bayésienne limitant le nombre d'effets non nuls (BayesB)

Une méthode approchée rapide : GSE, 2009

Estimation à partir d'une population de référence (training population) avec génotype et phénotype

Une multiplicité d'approches

Pas de méthode universelle

Beaucoup de travaux en cours

Approche INRA pour l'instant basée sur la SAM :

- beaucoup de QTL (20-40 par caractère),
- finement cartographiés,
- expliquant globalement une forte proportion de la V_g

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Quelle différence entre SAM et SG ?

- Nombre de QTL pris en compte
- Ciblage des QTL dans la SAM, sans a priori en SG
- Souvent SNP seuls en SG, haplotypes en SAM

MAIS Convergence forte si :

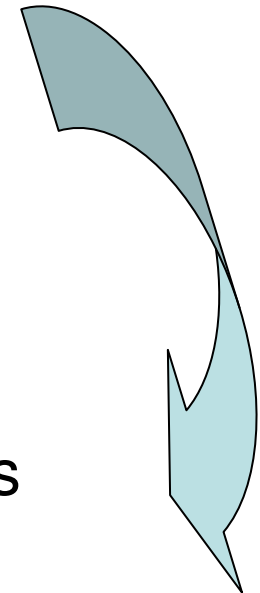
- la SAM inclut beaucoup de QTL
- la SG vérifie a posteriori les QTL utilisés
- la SG travaille avec des haplotypes

Apport en sélection

- **Prédiction directe de la valeur génétique des animaux, en fonction du génotype aux marqueurs**
- **Ne nécessite pas la mesure du phénotype chez le candidat ou ses apparentés proches**
- **Evaluation précoce => intervalle de génération**
- **Précision relativement élevée ($R^2 = 0,5$ à $0,7$ aujourd'hui)**
- **Précision homogène entre individus, relativement homogène (sous certaines conditions) entre caractères**

Comment ça marche ?

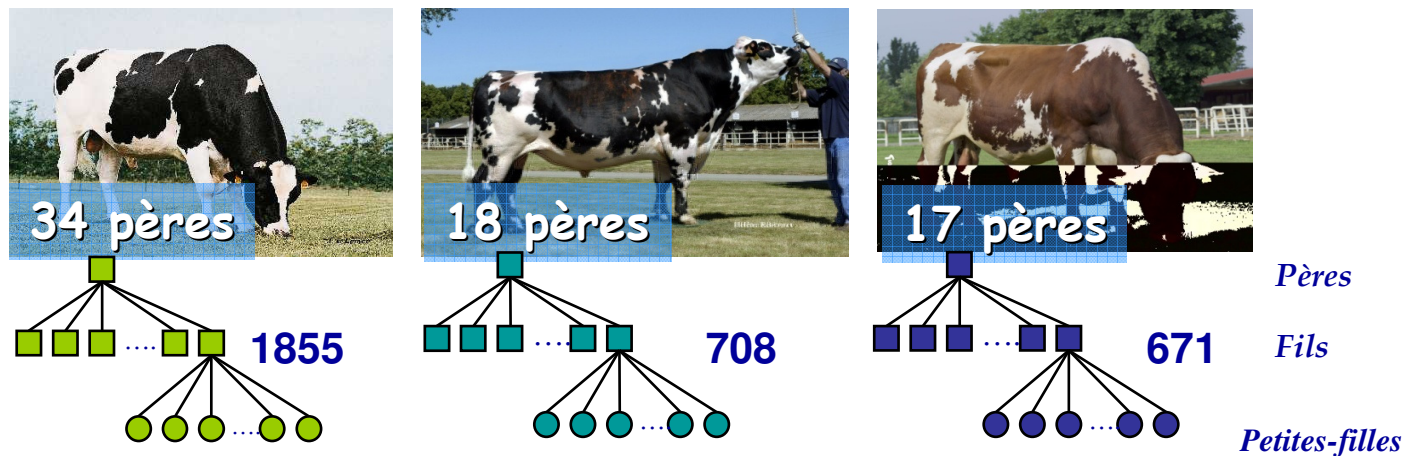
- **Population de référence :**
 - Population avec phénotypes et génotypes
 - Estimation des effets des marqueurs
 - L'efficacité dépend de la taille et de l' h^2
- **Population des candidats à la sélection**
 - Population présentant les mêmes associations
 - Génotypes
 - Prédiction de la valeur à partir des effets estimés



Quelle taille pour la population de référence ?

Quel renouvellement de la population de référence ?

Le programme Cartofine



- 3200 taureaux génotypés avec la puce 54k SNP
- testage sur descendance (100 filles avec phénotype)
- 25 caractères : production et composition du lait, fertilité, résistance aux mammites, facilité de naissance-mortinatalité, vitesse de traite, morphologie mamelle, corps, membres, ...

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Analyse LDLA (Meuwissen et al, 2000)

- Linkage Disequilibrium and Linkage Analysis
- Linkage analysis = analyse de liaison intra famille (de père)
=> puissance, robustesse, mais localisation peu précise
- LD analysis = analyse d'association, utilisant toute l'information (surtout maternelle)
=> puissance, manque de robustesse, mais localisation très précise
- LDLA combine les deux approches, assurant robustesse et précision de localisation

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Méthode LDLA

- Proba IBD entre QTL d'individus, y compris des fondateurs
- Haplotypes de 10 marqueurs (=> nécessité de reconstruire les phases)
- Genome scan complet, par pas de 2 SNP
- Clustering des haplotypes ayant une forte proba IBD
- Analyse REML (logiciel AI-REML interne), mono ou multi-QTL
- Test par rapport de vraisemblance

Marqueurs utilisés

- 38885 à 40757 SNP utilisés par race
- Carte précise : nécessaire pour travailler en haplotype
- MAF>5% dans l'échantillon de la race concernée

=> 15 SNP informatifs par cM

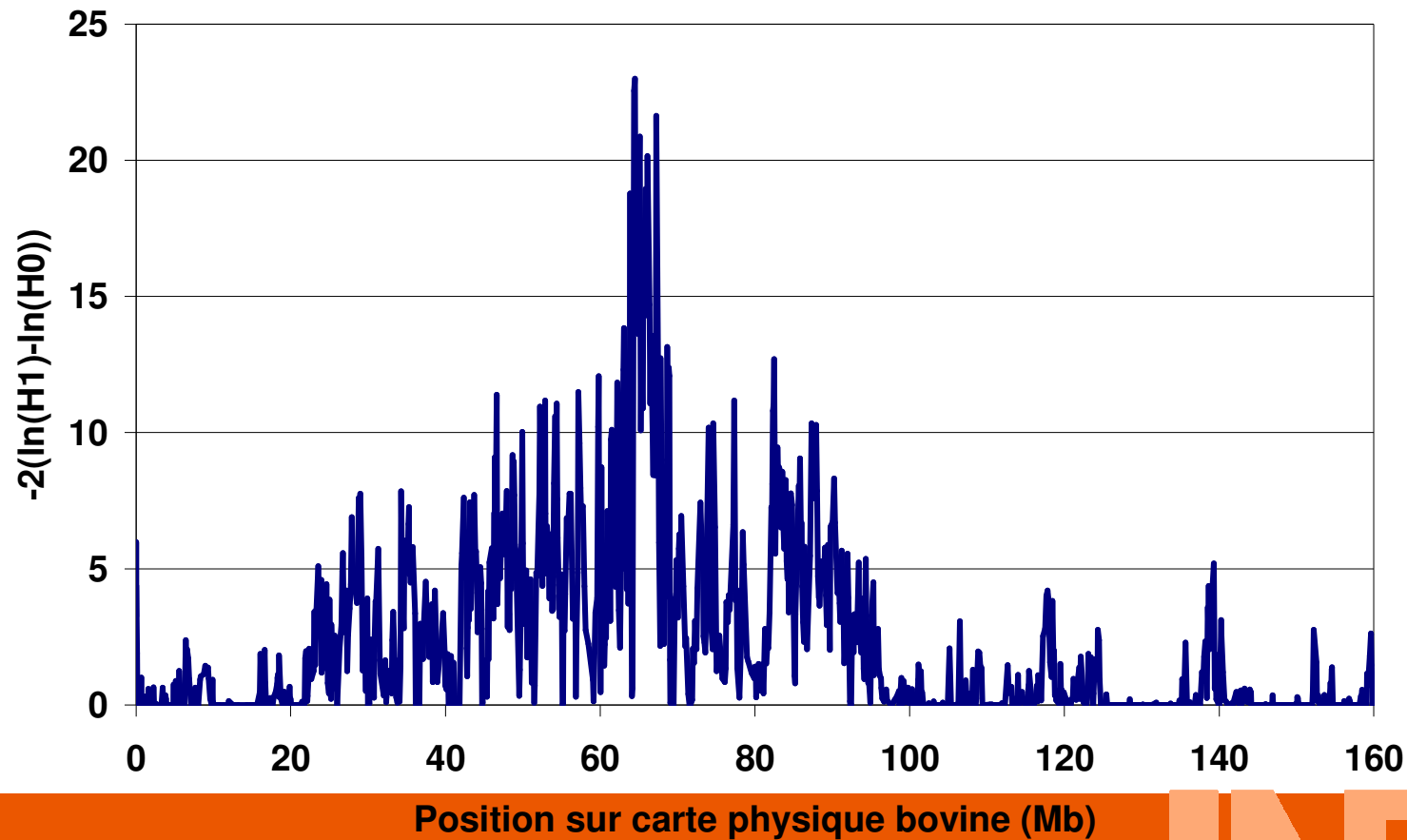
12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Des QTL finement cartographiés

Un exemple de profil : Lait, chrom 1, race Montbéliarde

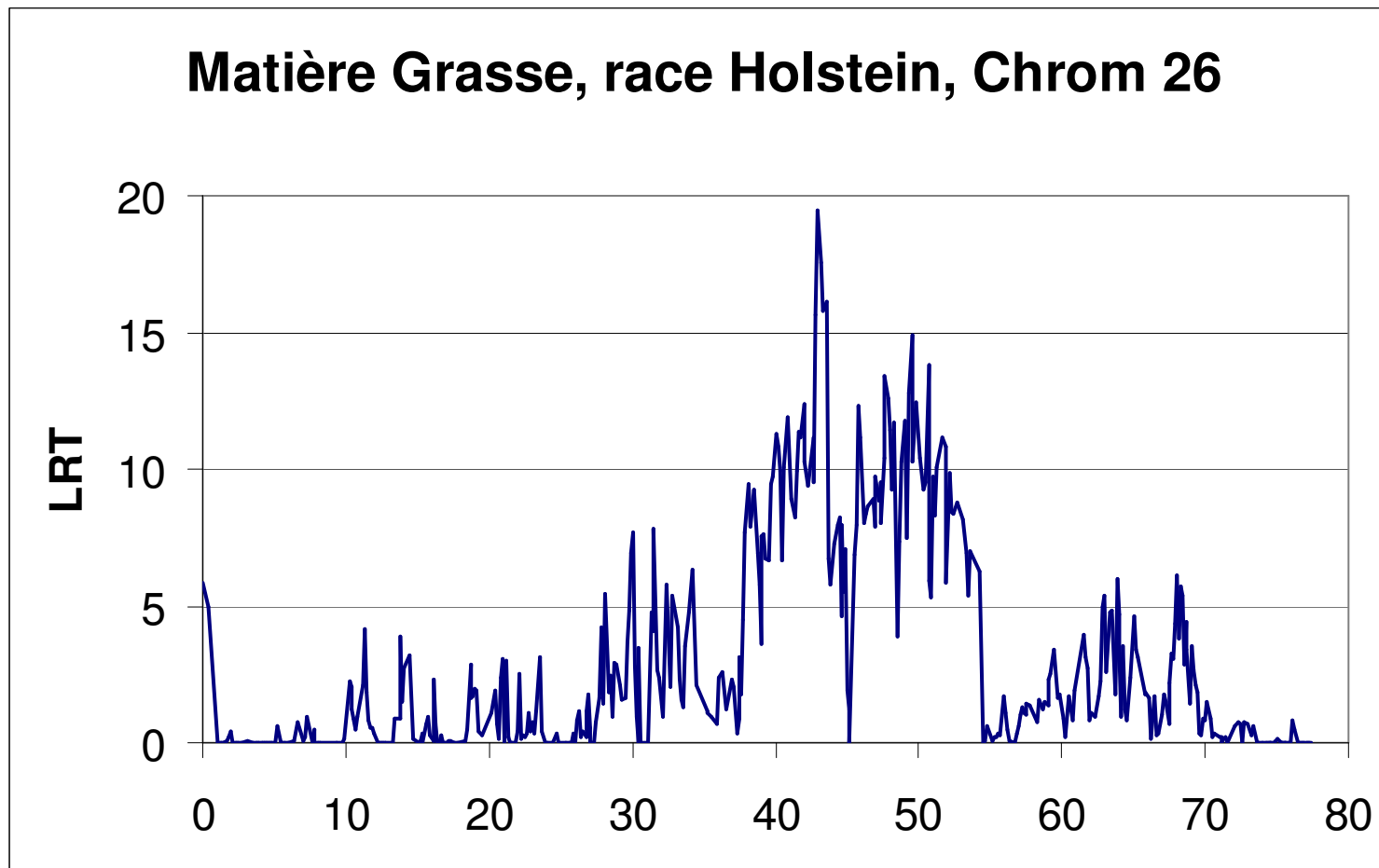


12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



Des QTL finement cartographiés



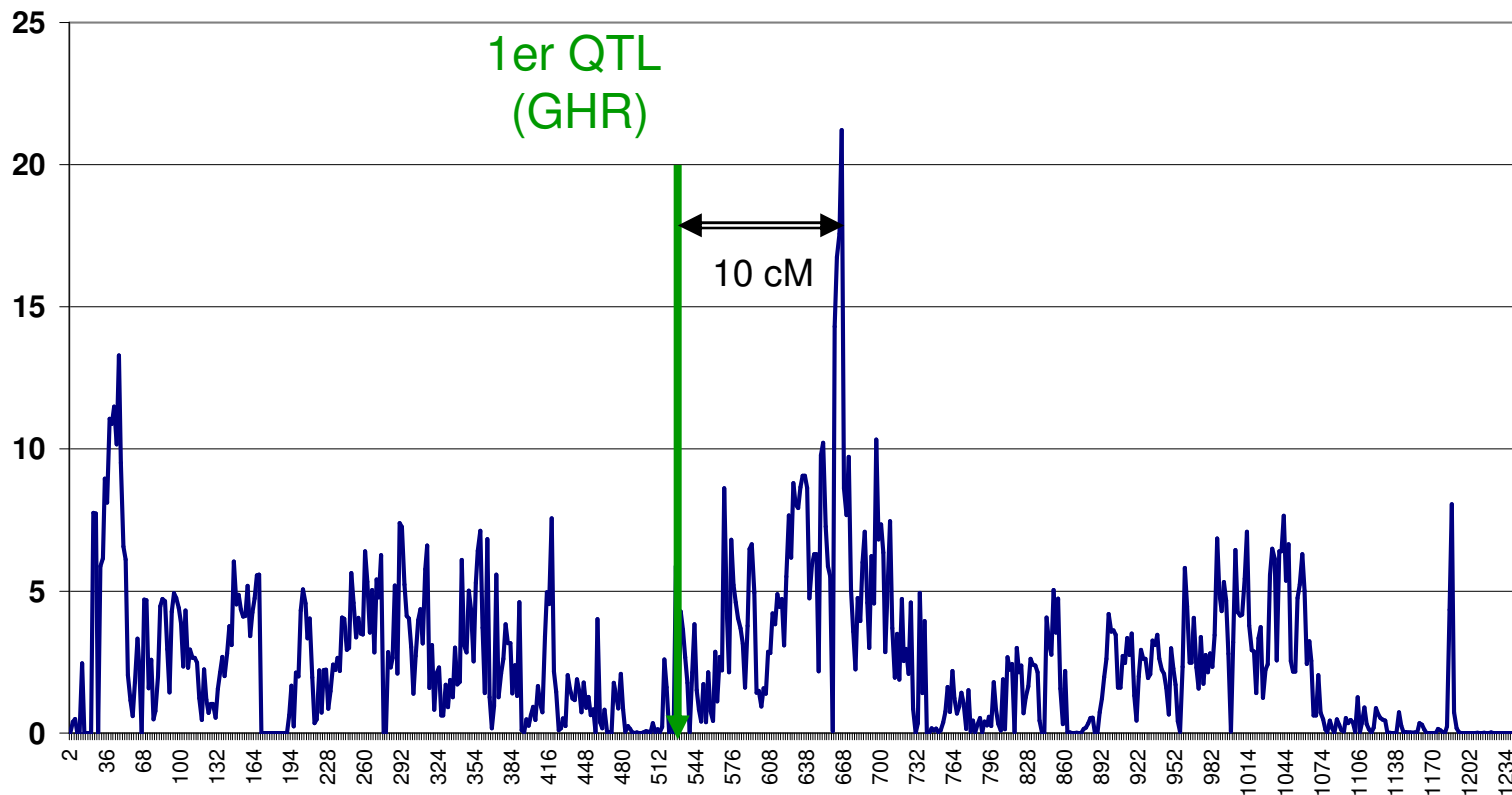
12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Une bonne séparation des QTL liés

Taux protéique, Chromosome 20, race Holstein



12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Bilan des résultats de Cartographie

- Le dispositif (avec des index sur descendance) et le DL apportent une très forte puissance
- Plusieurs dizaines (~50) de QTL détectés par caractère
- Un recouvrement très partiel entre races (un tiers)
- Une localisation très fine (souvent IC < 1 cM)
- Analyse multi-QTL efficace pour distinguer des QTL liés
- Plus de 50% de part de variance génétique expliqués par caractère
- 80% des QTL expliquent une part de variance inférieure à 6%, rares exceptions au-delà de 10%

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

SAM



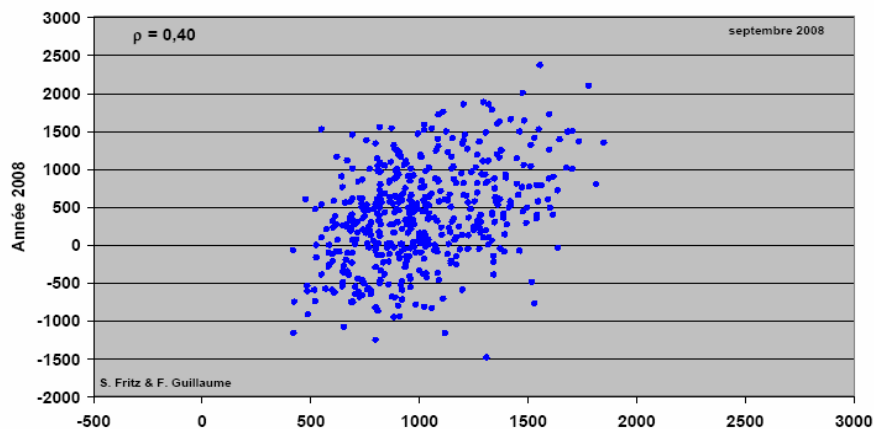
- 20 000 animaux génotypés par an à Labogena
- BLUP monocaractère, multiQTL (20-30), avec DL entre fondateurs, avec des haplotypes de 6 marqueurs

$$g_i = u_i + \sum_k (q_{ik1} + q_{ik2})$$

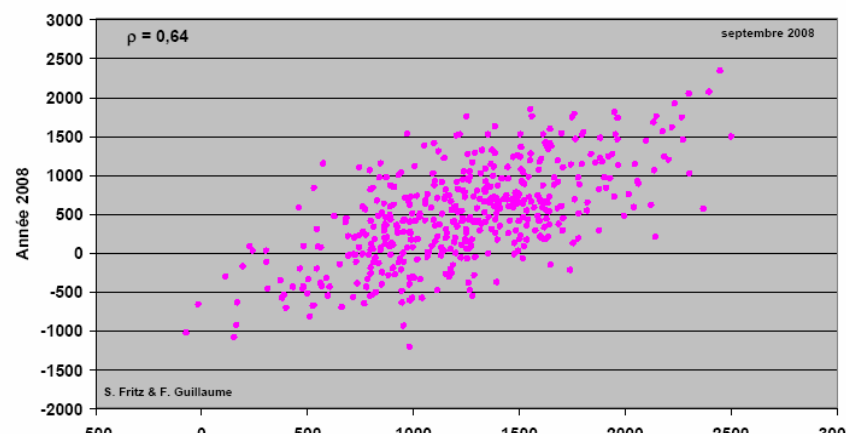
- Clustering selon les probabilités IBD pour limiter le nombre d'haplotypes à 20-35
- Précision estimée à partir des termes de l'inverse de la matrice du système d'équations

Une prédiction efficace de la valeur génétique

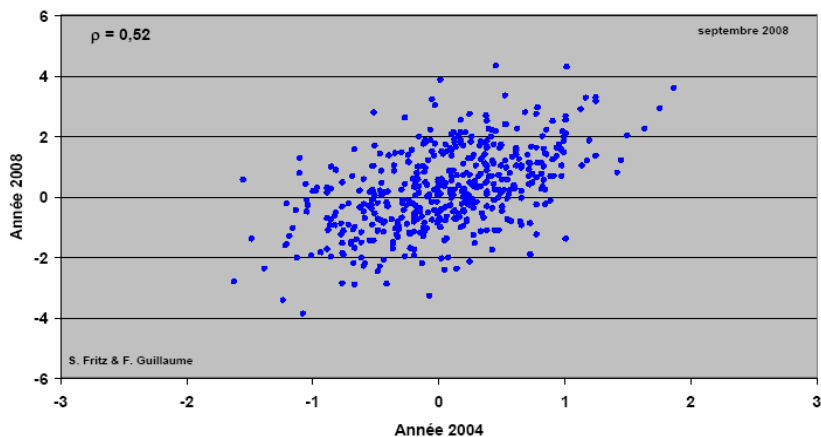
Modèle classique - LAIT



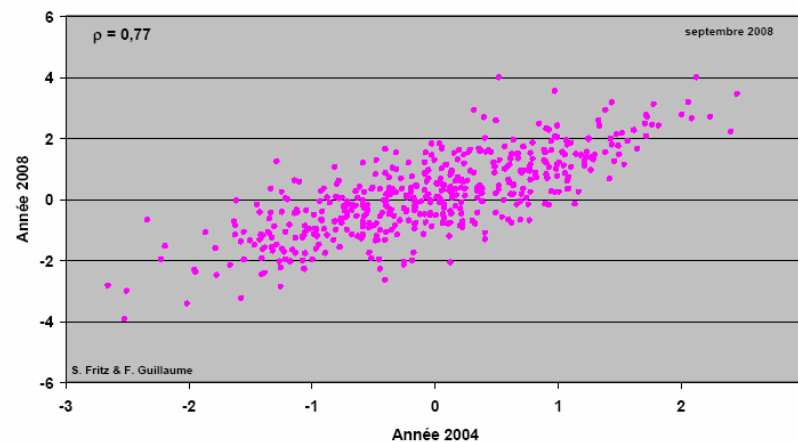
Modèle SAM2 - LAIT



Modèle classique TP



Modèle SAM2 - TP



12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Travaux en cours

- **Augmentation de la taille de la population de référence**
- **Comparaison des méthodes de SG (pgm ANR Amasgene)**
- **Analyse globale entre populations**
 - **Intra race : segments conservés de 300-500 kb**
 - **Entre races : segments conservés de 10-20 kb**
 - **Pour une densité actuelle de 1 SNP / 50 kb**
 - ⇒ **Développement actuel d'une puce de 660 kSNP**
- **Projet de caractérisation des QTL**

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Quelle amélioration génétique ?

- progrès fortement augmenté (x 2 ?), plus équilibré entre caractères
- arrêt du testage sur descendance
- équilibre organisationnel subtil à trouver entre génomique et phénotypage/contrôle de performances
- application dans d'autres espèces (valeur individuelle % coût du génotypage, croisement, politique...)
- puces réduites pour une sélection de masse

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

Collaborations

Financement : ANR, ApisGene, Inra, Unceia, Entreprises de sélection

Partenaires

**INRA : D Boichard, T Druet, J Tarres, V Ducrocq, JJ Colleau, M Gautier,
S Ben Jemaa, M Boussaha, C Grohs, A Eggen,**

CNG : I Gut, A Boland, D Zelenika

UNCEIA : S Fritz, A Baur, A Malafosse, L Journaux

Institut de l'Elevage : F Guillaume

Labogena : MY Boscher, Y Amigues, L Genestout

12 mai 2009

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA