



Sélection assistée par marqueurs

*Didier BOICHARD, François GUILLAUME,
Joaquim TARRES, Aurélia BAUR, André EGGEN,
Tom DRUET, Sébastien FRITZ*

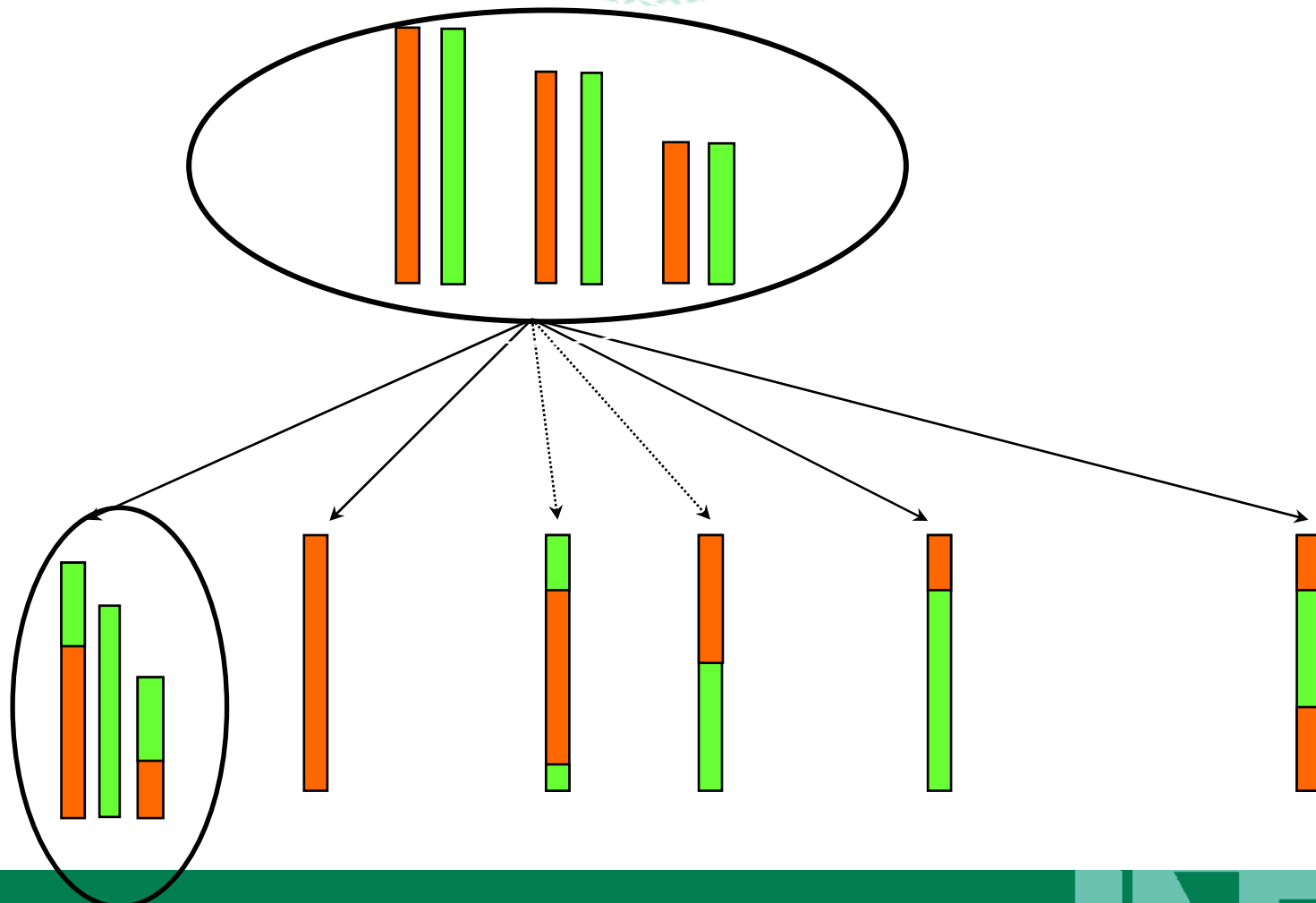
Agenciae 24 novembre 2008

1

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



Transmission des gènes entre générations



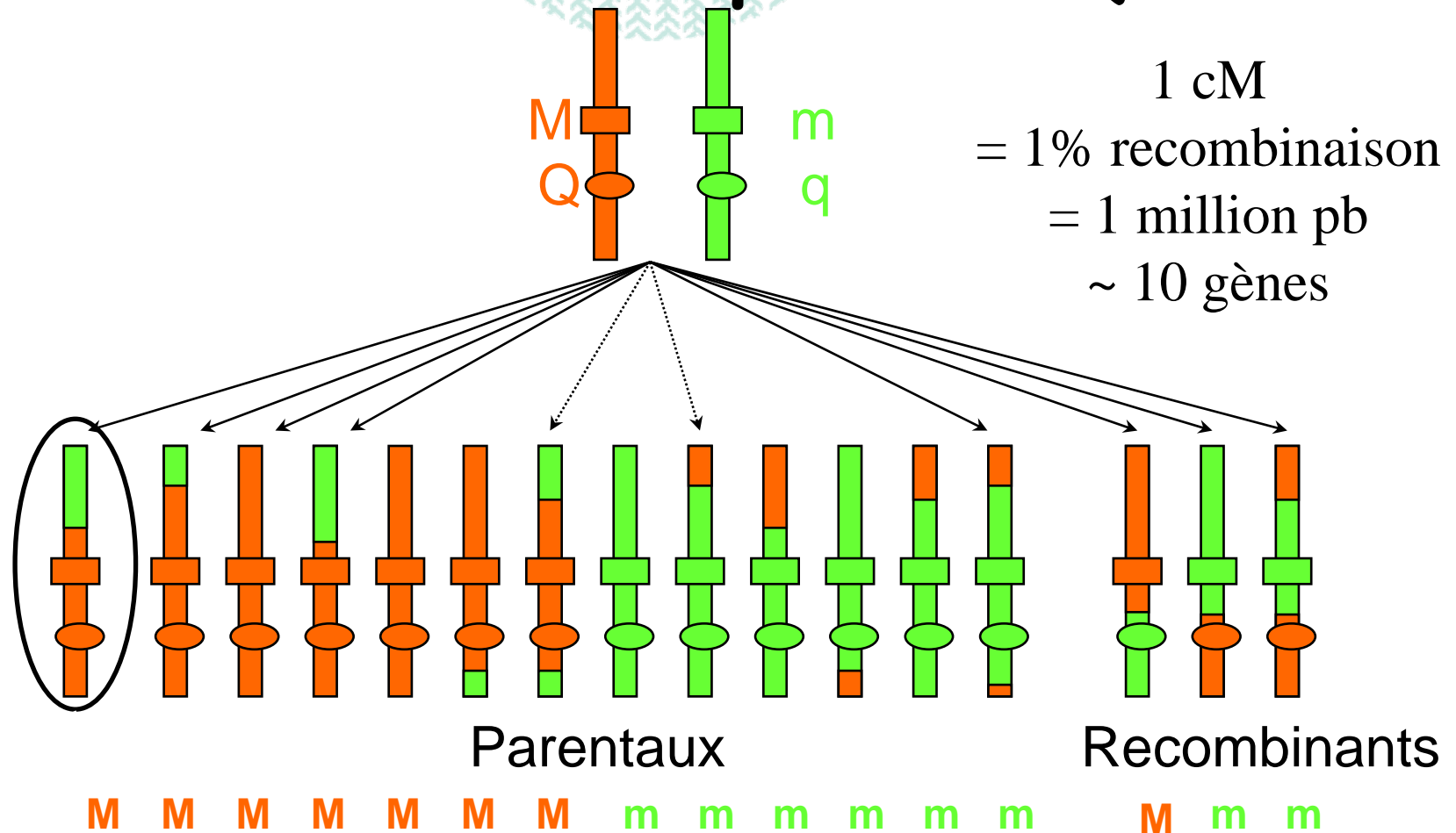
Agenciae 24 novembre 2008

2

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



Association marqueur - QTL



Propriétés des marqueurs

- **Le marqueur est « loin » du gène :**
 - Recombinaisons assez fréquentes
 - Association intra famille uniquement
 - Sélection intra famille
- **Le marqueur est très proche du gène**
 - Recombinaisons rares
 - Association générale à l'échelle de la population
 - Déséquilibre de liaison populationnel
 - Sélection à l'échelle de la population

Evolution des marqueurs

- **Jusqu'à récemment :**
 - Quelques centaines de marqueurs microsatellites
 - Statistiquement, plusieurs cM entre le QTL et les marqueurs les plus proches
- **Actuellement**
 - Des dizaines de milliers de marqueurs SNP
 - Des marqueurs à proximité immédiate du QTL
 - ⇒ Cartographie fine des QTL
 - ⇒ Evolution des dispositifs de cartographie
 - ⇒ SAM de nouvelle génération

L'état de l'art : Les puces à SNP

illumina
making sense out of life

LOG IN TO ICOM ORDER SEARCH

ABOUT ILLUMINA PRODUCTS & SERVICES SUPPORT COMMUNITY CORPORATE

products & services

- overview
- ▣ systems & software
- ▣ dna analysis solutions
 - overview
 - snp genotyping
 - whole-genome
 - human-1
 - humanhap240S
 - humanhap300-duo
 - humanhap300-duo+
 - humancnv370-duo
 - humanhap550
 - humanhap550+
 - human650Y
 - human1M
 - focused content
 - cancer snp panel
 - dna test panel
 - humanNS-12
 - mhc panel set
 - goldengate custom panels
 - iselect custom panels

whole-genome genotyping: human1M

The ability to genotype over one million SNPs has now become a reality. With the introduction of the Human1M BeadChip to the Infinium product family, researchers can interrogate over one million diverse genetic variants on a single array. The new Human1M BeadChip combines an unprecedented level of content for both whole-genome and copy number variation (CNV) analysis, along with additional unique, high-value genomic regions of interest. The Human1M BeadChip will be powered by Illumina's revolutionary Infinium Assay, providing industry-leading data quality, genomic coverage, and intelligent probe selection. This product is expected to enter the market by the close of the second quarter, 2007.

important information

- product literature
- publications
- faqs
- have a rep contact me



La puce SNP Illumina bovine

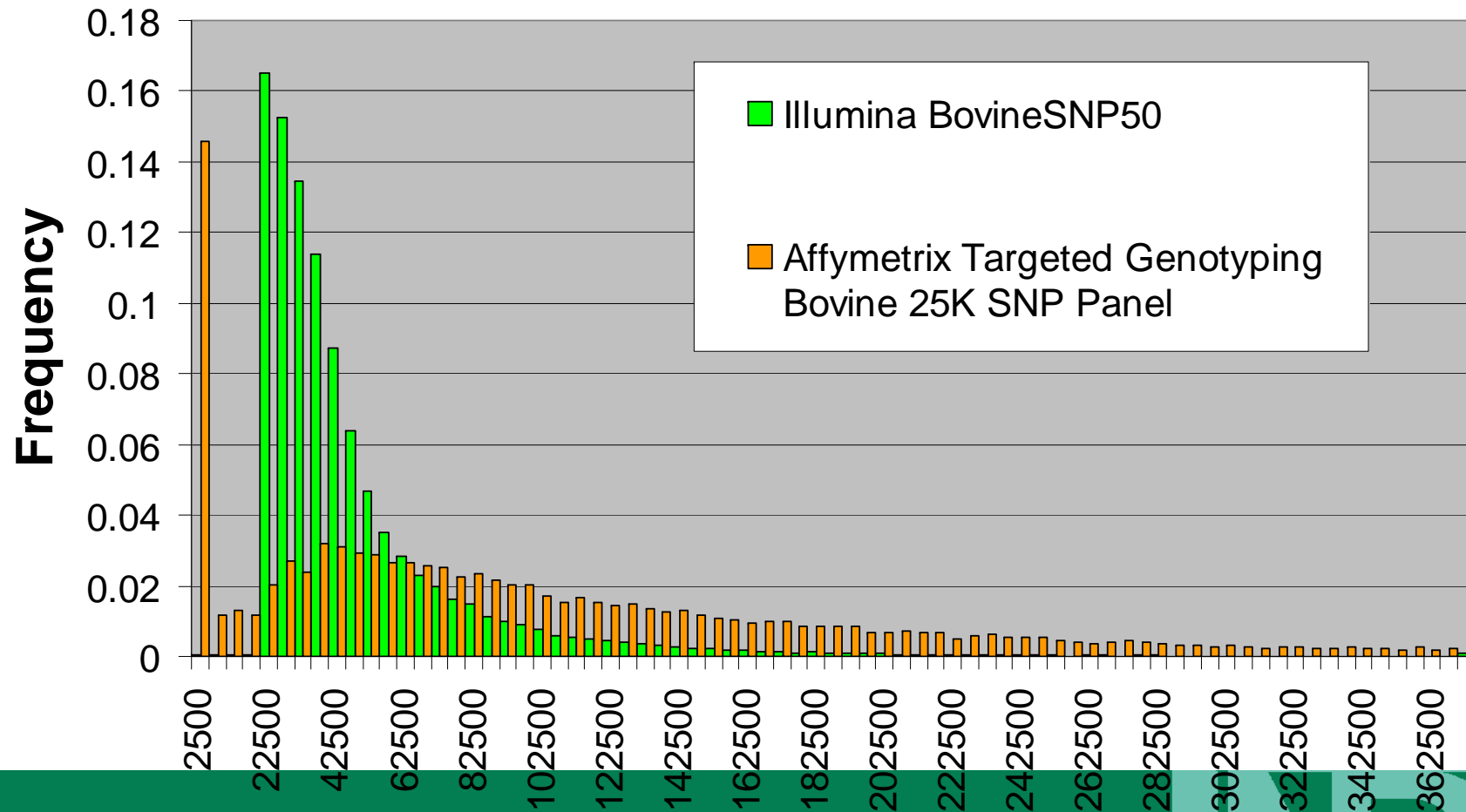
- 54 000 SNP répartis de façon homogène sur l'ensemble du génome bovin
- A un coût très raisonnable

FIGURE 1: BOVINESNP50 BEADCHIP



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

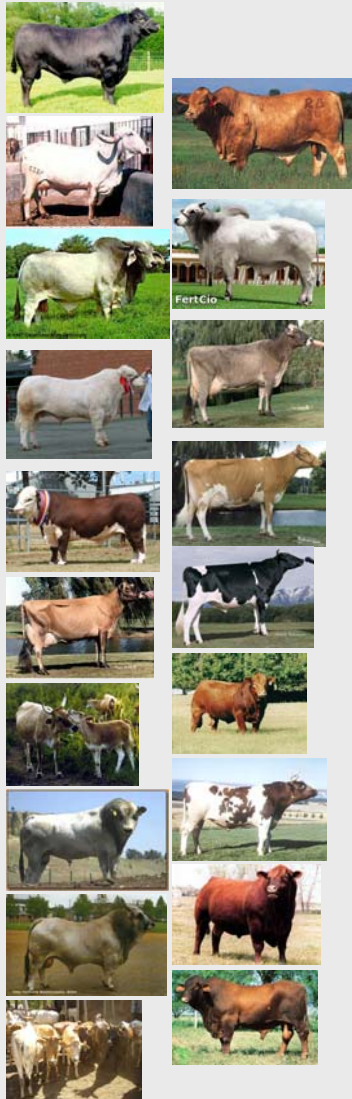
Distribution des intervalles entre SNP



A *Illumina Bovine SNP50*
Mean Interval : 45,878 Kb

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT





Ageneae 24 novembr

9

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**

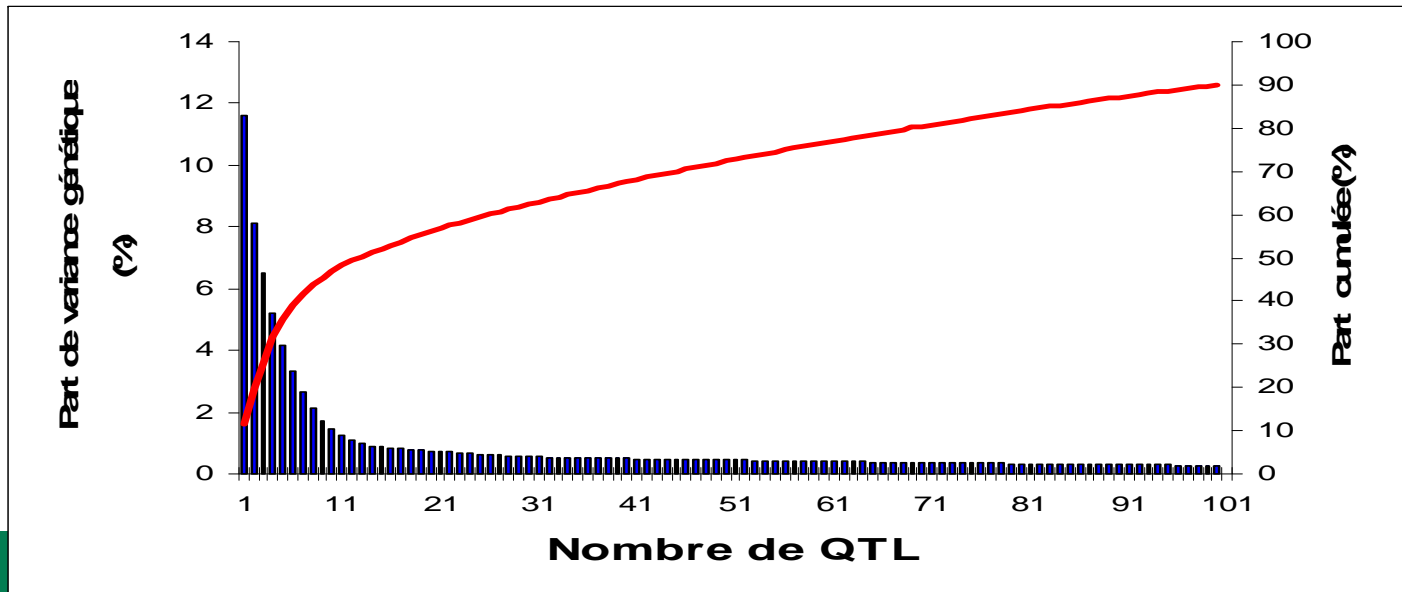


Puces commerciales disponibles

- Homme : 1 million et 384k
 - Bovin : Affymetrix 10k en 2005, Illumina 54k en 2007, plus forte densité en préparation
 - Cheval, Mouton, Porc : 50-60k
 - Poule, Souris prochainement...
- Prix catalogue : 200-250 euros, grosses remises sur volume

Le déterminisme génétique des caractères

- distribution « en L »
 - Peu de gènes à gros effet
 - Plusieurs gènes à effet intermédiaire (qq % V_g)
 - Beaucoup de gènes à petit effet

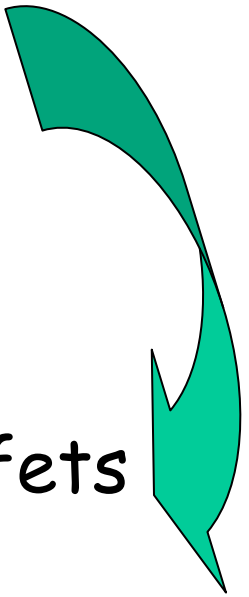


Sélection assistée par marqueurs

Un constat : on connaît très peu de mutations causales

- Sélection sur les principaux QTL (1 à quelques dizaines)
- Les QTL sont localisés, mais non identifiés
- On sélectionne sur des marqueurs proches
- Mais on ne sélectionne pas le reste du génome
- SAM1 et SAM2 (marqueurs très proches, DL)

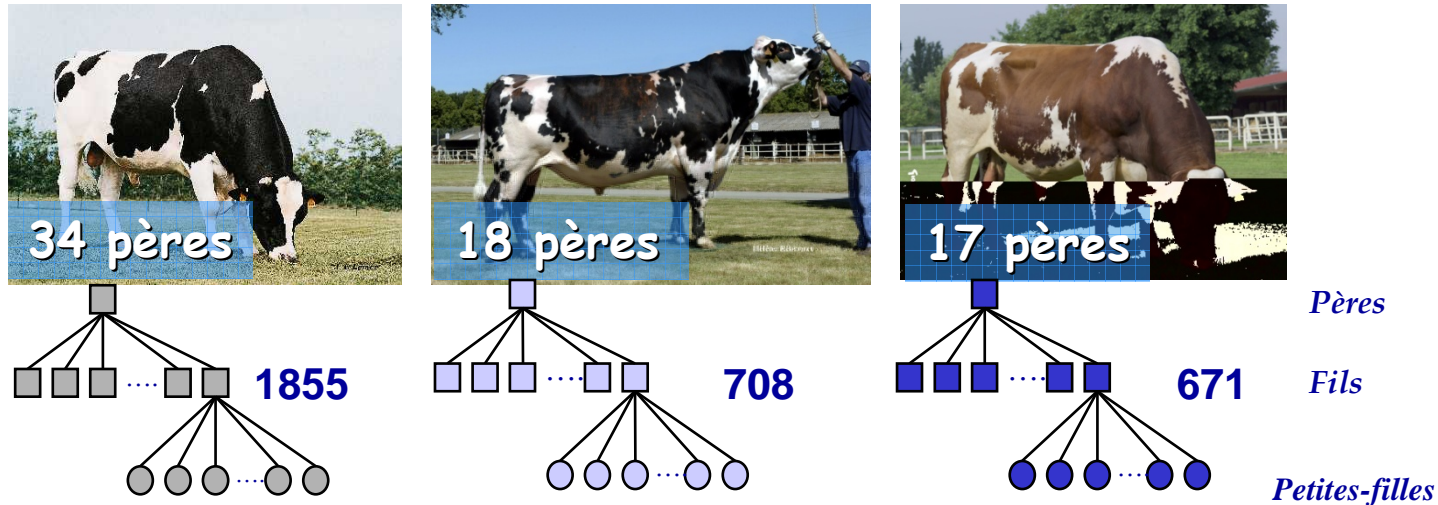
Comment ça marche ?

- **Population de référence :**
 - Population avec phénotypes et génotypes
 - Estimation des effets des marqueurs
 - L'efficacité dépend de la taille !
 - **Population des candidats à la sélection**
 - Génotypes
 - Prédiction de la valeur à partir des effets estimés
- 

Apports de la SAM par rapport à la sélection classique

- Ne nécessite pas de phénotype chez le candidat (caractères non mesurables, invasifs, tardifs...)
- Efficace même si h^2 basse, si les effets sont bien estimés
- Progrès mieux équilibré sur les différents caractères de l'objectif de sélection
- Estimation précoce => diminution des coûts, diminution de l'intervalle de génération, augmentation de l'intensité de sélection
- Permet de prendre en compte des caractères nouveaux

Le programme ANR Cartofine



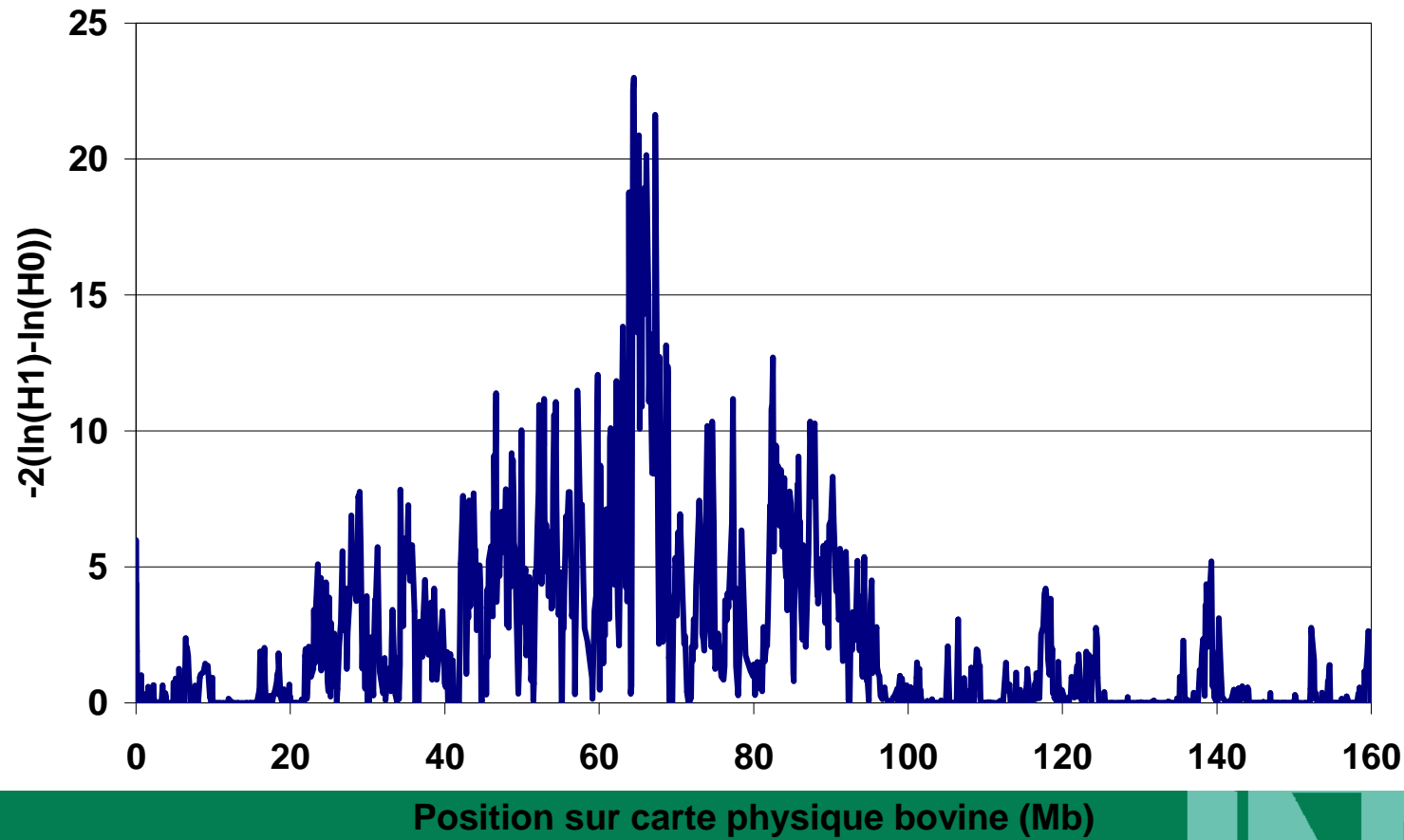
- ~3200 taureaux génotypés avec la puce 54k SNP
- testage sur descendance (100 filles avec phénotype)
- 25 caractères : production et composition du lait, fertilité, résistance aux mammites, facilité de naissance-mortinatalité, vitesse de traite, morphologie mamelle, corps, membres, ...

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

Résultats

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



Agencae 24 novembre 2008
17

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



Bilan des résultats

- Plusieurs dizaines de QTL détectés par caractère
- Holstein : 342 QTL,
Normande : 282 QTL,
Montbéliarde : 301 QTL
(dont environ 10% faux positifs...)
- Une localisation très fine (souvent IC < 1-2 cM)
- Plus de 50% de part de variance génétique expliqués par caractère

Mise en place de la SAM de 2ème génération

- SAM de seconde génération, basée sur des SNP et une cartographie fine des QTL
- Prédiction précoce de la valeur des jeunes reproducteurs potentiels, à partir de leur génotype aux marqueurs
- 10 000 animaux génotypés par an (Labogena)
- Méthode actuelle de prédiction (octobre 2008) : BLUP monocaractère, multiQTL (jusqu'à 25), avec DL entre fondateurs, avec des haplotypes de 6 à 10 marqueurs et clustering

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

- Comparaison avec la sélection génomique courant 2009

Méthode

- Population : génotypés + 2 générations d'ancêtres
- Phénotype = sous produit de l'évaluation génétique polygénique (DYD et YD)

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

- valeur polygénique, $\text{var}(u) = A \sigma_g^2$
- Valeur qtl : $\text{var}(q) = G \sigma_q^2$
- G constitué des probabilités IBD au QTL, conditionnellement à l'haplotype de 6 à 10 marqueurs (Meuwissen et al, 2000)
- Nécessite la reconstruction préalable des phases
 $m1 \ m2 \ n1 \ n2 \Rightarrow (m1 \ n1) \ (m2 \ n2) \ \text{ou} \ (m1 \ n2) \ (m2 \ n1)$
- Clustering : regroupement des haplotypes sur la base de leur proba IBD

Méthode (exemple pour 2 QTL q_1 et q_2)

$$y = 1'\mu + X_1 q_1 + X_2 q_2 + Z u + e,$$

$$\text{var}(q_1) = G_1 \sigma_{q_1}^2, \text{var}(q_2) = G_2 \sigma_{q_2}^2, \text{var}(u) = A \sigma_g^2, \text{var}(e) = W^{-1} \sigma_e^2$$

NB : réduction de q et G , après clustering

$$\begin{bmatrix} 1'W1 & 1'WX_1 & 1'WX_1 & 1'WZ \\ X_1'W1 & X_1'WX_1 + G_1^{-1}\lambda_1 & X_1'WX_2 & X_1'WZ \\ X_2'W1 & X_2'WX_1 & X_2'WX_2 + G_2^{-1}\lambda_2 & X_2'WZ \\ Z'W1 & Z'WX_1 & Z'WX_2 & Z'WZ + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{q}_1 \\ \hat{q}_2 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 1'Wy \\ X_1'Wy \\ X_2'Wy \\ Z'Wy \end{bmatrix}$$

Précision de g déduite de l'inverse de 

Validation de la SAM2

~900 Candidats mis au testage, recevant leurs premières estimations de valeur sur descendance en 2008

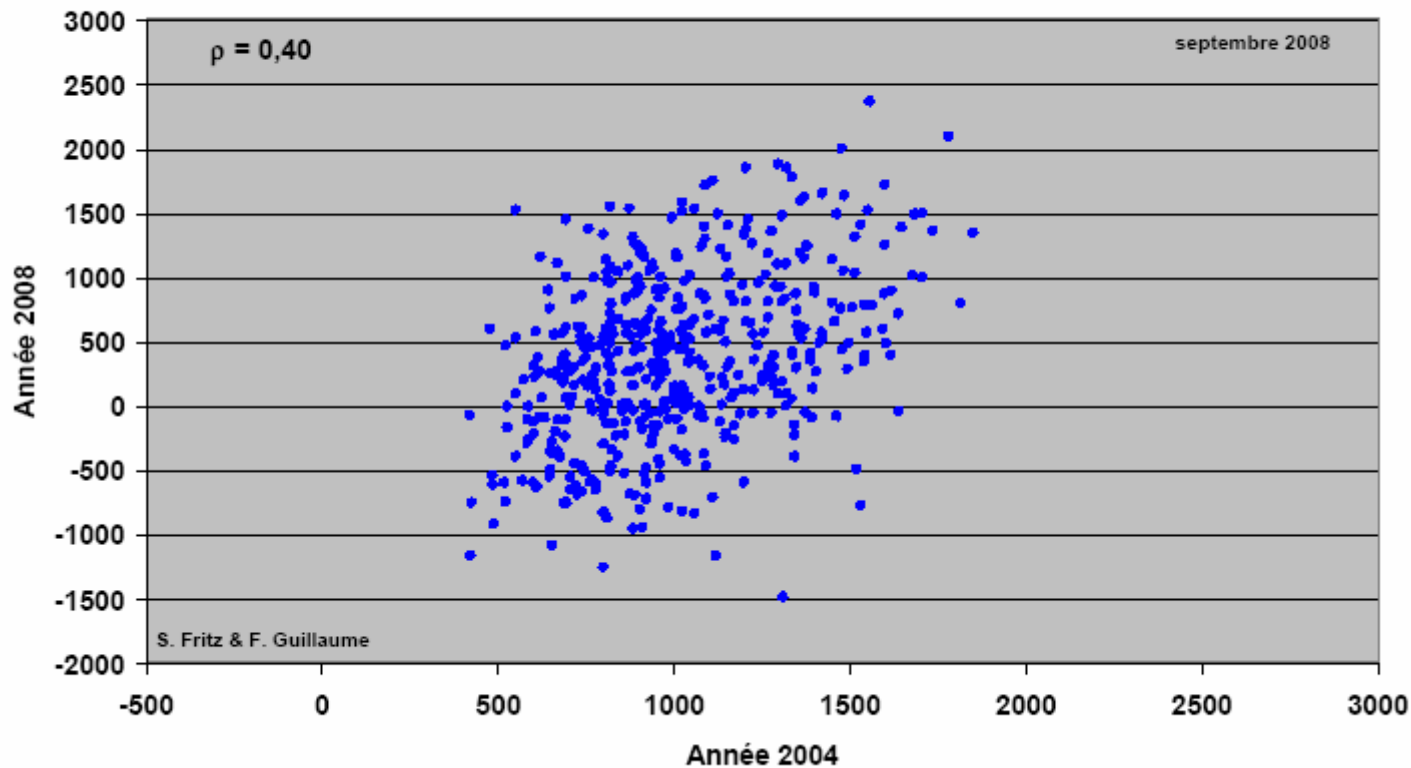
Prédiction de leur valeur sans leur phénotype

Comparaison * Descendance vs Ascendance
 * Descendance vs SAM2

Lait prim'Holstein

Cas polygénique

Modèle classique - LAIT



Corrélation = 0.40 \Rightarrow $CD_{2004} < 0,30$

$h^2=0.30; N=468$

Agencae 24 novembre 2008
23

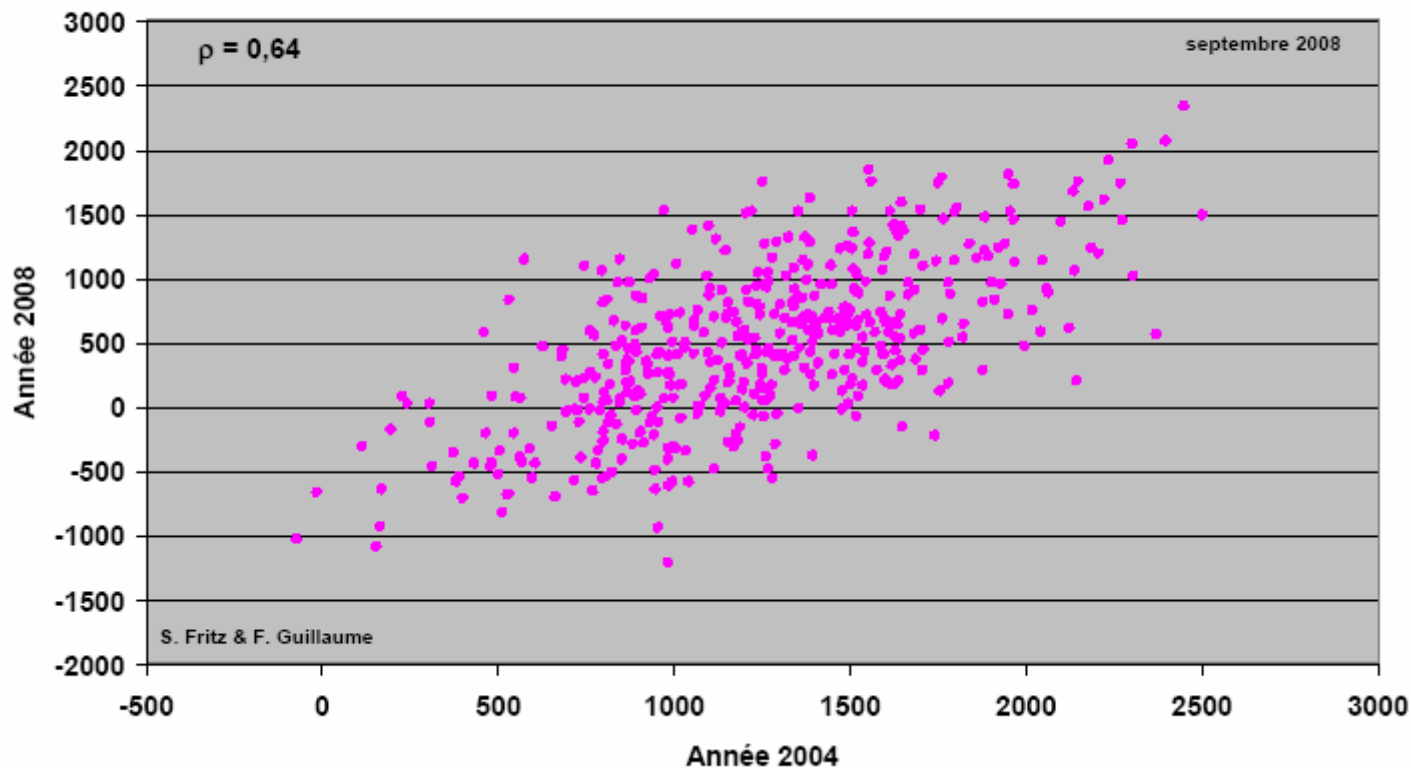
ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



Lait prim'Holstein

Cas SAM2

Modèle SAM2 - LAIT



Corrélation = 0.64 \Rightarrow $CD_{2004} > 0,50$

$h^2=0.30$; $N=468$

Agencae 24 novembre 2008
24

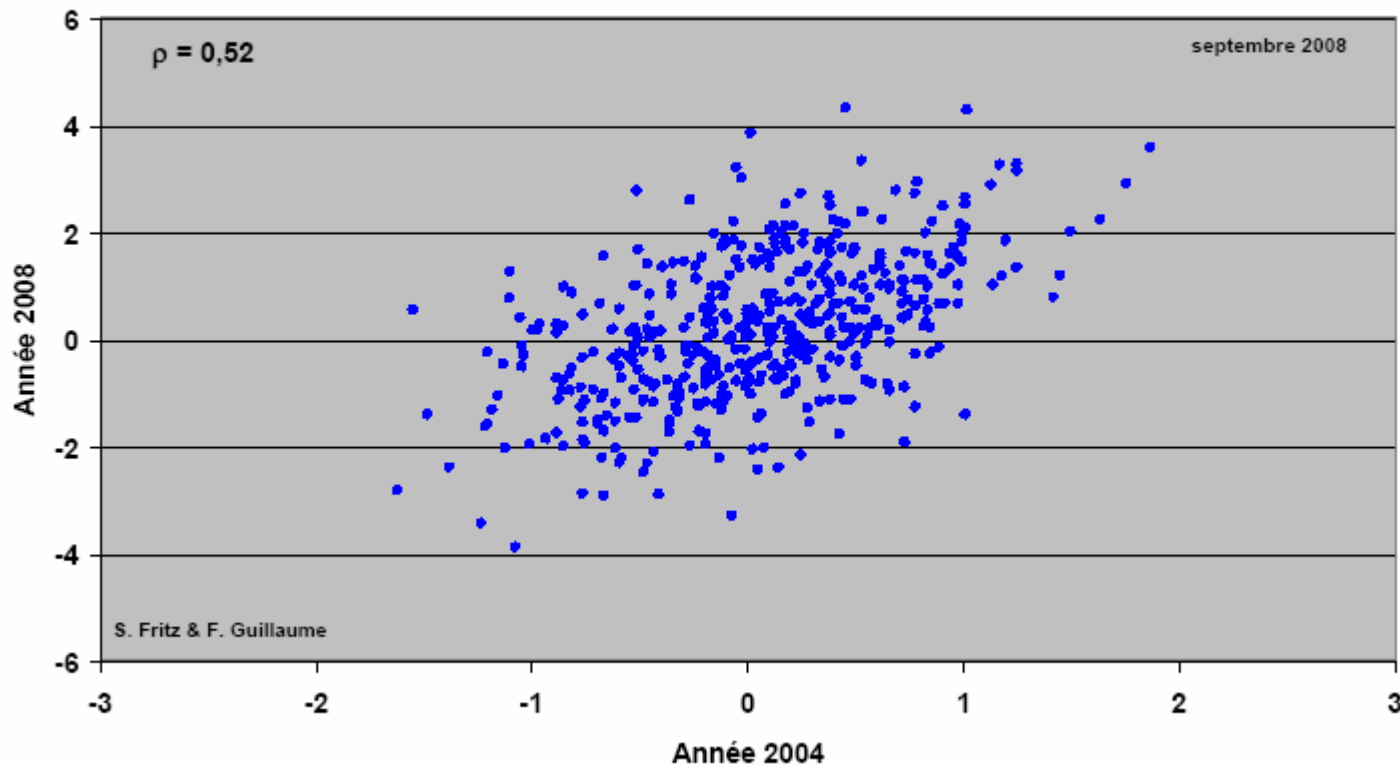
ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



TP prim'Holstein

Cas polygénique

Modèle classique TP



Corrélation = 0.52 \Rightarrow $CD_{2004} \sim 0,35$

$h^2=0.50$; $N=468$

Agenciae 24 novembre 2008
25

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



TP prim'Holstein Cas SAM2

Modèle SAM2-TP



Corrélation = 0.77 \Rightarrow $CD_{2004} > 0,70$

$h^2=0.50$; $N=468$

Agencae 24 novembre 2008
26

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



Quelle différence entre MAS 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 entre SAM et SG si :

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

Conclusions



- Le programme SAM2 est opérationnel depuis octobre 2008 et remplace le programme SAM1 (2000-2008)
- Il sera comparé à la SG en 2009.
- Le programme SAM utilise un outil de génotypage, la puce 54k d'Illumina, qui s'avère d'excellente qualité => le nombre d'individus typés va croître très rapidement
- La précision de localisation va s'accroître avec la taille de la population typée, facilitant également la caractérisation des QTL
- Les conséquences en sélection des bovins laitiers sont énormes : sans doute l'arrêt du testage sur descendance et une réforme importante de la diffusion du progrès génétique.
- Synergie entre caractérisation de QTL et sélection

Collaborations

Financement : ANR, ApisGene, Inra, Entreprises de sélection

Partenaires

INRA-SGQA : D Boichard, T Druet, J Tarres

INRA-LGBC : A Eggen, M Gautier, S Ben Jemaa, M Boussaha, C Grohs

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

8 entreprises de sélection : CREA VIA, GDO, AMELIS, DYNAM'IS,
MIDATEST, GNA, UMOTEST et Jura Bétail