

# L'utilisation du DL en populations croisées

# Problématique

Le DL entre un marqueur et un QTL dans une population croisée a trois origines:

- le DL historique de la population 1
- le DL historique de la population 2
- le DL créé par le croisement

# Le DL dans les populations croisées

L'effet du SNP peut être spécifique de la race.

Pour accéder à un marqueur en DL avec une mutation ancestrale, on peut considérer :

- Des marqueurs denses
- Des races croisées proches

# Les travaux récents en SG ou détection de QTL

- Prédiction de valeurs génétiques en race pure à partir de populations croisées (Dekkers, 2007; Ibanez et al, 2009; Toosi et al, 2010)
- Prédiction de valeurs génétiques en race pure à partir de plusieurs races pures (Roos et al, 2009; Hayes et al, 2009; Kizilkaya et al, 2010)

# Prédiction des GBV en race pure à partir de populations croisées

- Ibanez- Escriche, Fernando, Toosi, Dekkers, 2009. Genomic selection of purebreds for crossbred performance. GSE, 41, 12.
- Toosi, Fernando, Dekkers, 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci., 88,32-46

# Modèle BSAM(breed specific SNP allele model) de Ibanez et al (2009)

- HYPOTHESE: L'origine raciale de l'allèle est connue sans erreur.

- **Modèle classique:** 
$$y_i = \mu + \sum_j X_{ij} \beta_j \delta_j + e_i, \quad (1)$$

Ou  $X_{ij}=0,1$  ou 2 en fonction du génotype  $i$  au marqueur  $j$ ,  $\beta_j$  effet du marqueur  $j$ ,  $\delta_j$  indicateur 0/1 de la présence ou non dans le modèle du marqueur  $j$ .

- **Modèle BSAM:** 
$$y_i = \mu + \sum_j (A_{ijk}^S \beta_{jk}^S \delta_j^S + A_{ijl}^D \beta_{jl}^D \delta_j^D) + e_i, \quad (2)$$

Ou  $A_{ijk}^S=0$  ou 1 en fonction de l'allèle SNP reçu du père d'origine raciale  $k$  au locus  $j$ ,  $\beta_{jk}^S$  effet de l'allèle du père,  $\delta_j^S$  est un indicateur 0/1 de la présence ou non dans le modèle du marqueur portant l'allèle du père

## 2 fois plus de paramètres à estimer avec BSAM

- La méthode Bayes B est utilisées pour estimer les paramètres du modèle

# Description des Simulations

	<b>Ibanez et al, 2009</b>	<b>Toosi et al, 2010</b>
Nb de générations de divergence	50 ou 550 ou indépendante	25, 50
Nb marqueurs	500 ou 2000 SNP/1 M	500, 1000, 2000, 40000/ 1M
Nb QTL	<b>30 QTL</b> dans une distribution gamma (0.4, 1/1.66)	<b>100 QTL</b> dans une distribution gamma (0.4, 1/1.66)
Taille de la population	1000 générations Ne=500 50 générations Ne=100 Puis 1000 ou 4000	1000 générations Ne=500 50 générations Ne=100 Puis 1000
H <sup>2</sup>	0.3	0.3
Nb réplifications	40	160
Croisement entre races	AB, AC, (AB) <sup>2</sup> , (AB)C, (AB)(CD), (Adm_AB)B, Adm_ABCD)B	AB, (AB)C, (AB)(CD)
model	Tous les marqueurs ensembles	Tous les marqueurs ensembles, <b>effet race ou pas (BSAM)</b>
méthode	Bayes B, prior Prob=5%	Bayes B, prior Prob=5%

# Pot pourri des conclusions

- BSAM décevant: un peu meilleur que le modèle classique lorsque :
  - grande pop de référence (4000)
  - races très distantes (500 générations ou indépendante)
  - peu de marqueurs (500/1M)
- 2 fois plus de paramètres à estimer
- Avec un modèle classique et des races proches (50 générations) → précision équivalente en race pure ou croisée
- Précision en race pure ↓ avec le nombre de races croisées
- Précision de GBV d'une race pure : 4000 animaux croisés de 4 races (50 générations de divergence)  $\equiv$  1000 animaux race pure.
- Dans les mêmes conditions : précision Ibanez > Toosi de 6% car moins de QTL simulés
- Précision ↓ de 25% si 10chr au lieu de 1 chr (1000 QTL au lieu de 100)
- Avec des puces très denses (1 000 000 SNP), la perte de précision est au maximum de 8% en races croisées/ race pure (si présence de la race pure dans la pop croisée sinon 47%)



# Prédiction des GBV en race pure à partir de plusieurs races pures

- Hayes, Bowman, Chamberlain, Verbyla, Goddard, 2009. Accuracy of genomic breeding values in multi breed dairy cattle populations. GSE, 41:51.  
→ Populations réelles bovines
- Kizilkaya, Fernando, Garrick, 2010. Genomic prediction of simulated multibreed and purebred performance using observed 50 000 SNP genotypes. J Anim Sci, 88:544-551.  
→ Simulation à partir de populations vraies
- Roos, Hayes, Goddard, 2009. Reliability of genomic predictions across multiple populations. Genetics, 183: 1545-1553.  
→ Populations simulées

# Mat. Meth: Hayes et al., 2010

- Populations vraies: Holstein (R=781+V=400) et Jersey (R=287+V=77)
- Realised accuracy= $r(\text{GEBV}, \text{ABV})$
- ABV= valeur génétique estimée sur descendance (80 filles/père)

Methodes: Gblup, BayesA, BayesSSVS

# Résultats: Hayes et al., 2010

**Table 3: Accuracies of GEBV using either GBLUP or SNP effects from BAYESA or BAYES\_SSVS to predict GEBV**

Reference set	Validation set	Method	Protein	Fat	Milk	Prot%	Fat%
Holstein	Holstein	GBLUP	0.49	0.44	0.59	0.61	0.62
		BAYESA	0.47	0.44	0.59	0.59	0.71
		BAYES_SSVS	0.47	0.44	0.59	0.58	0.70
	Jersey	GBLUP	-0.06	-0.02	-0.02	-0.06	0.23
		BAYESA	0.24	0.35	0.37	0.33	0.63
		BAYES_SSVS	0.27	0.31	0.23	0.29	0.42
Jersey	Holstein	GBLUP	0.03	-0.01	-0.01	0.03	0.11
		BAYESA	0.01	0.02	-0.02	0.05	0.17
		BAYES_SSVS	0.03	0.04	0.01	0.02	0.11
	Jersey	GBLUP	0.53	0.41	0.63	0.62	0.72
		BAYESA	0.43	0.37	0.59	0.51	0.67
		BAYES_SSVS	0.43	0.37	0.59	0.51	0.65
Holstein & Jersey	Holstein	GBLUP	0.49	0.45	0.59	0.61	0.62
		BAYESA	0.47	0.44	0.55	0.54	0.69
		BAYES_SSVS	0.46	0.45	0.55	0.54	0.70
	Jersey	GBLUP	0.53	0.42	0.56	0.60	0.73
		BAYESA	0.47	0.51	0.58	0.67	0.82
		BAYES_SSVS	0.47	0.51	0.58	0.67	0.82

# Mat. & Meth. : Roos et al 2010

- Simulation de 2 populations de 1000 ind.
- 6, 30 ou 300 générations de divergence  $\equiv$  divergence entre Aus holstein et Dutch Holstein ou NZ Holstein ou Au Angus
- 150 QTL (exponentiel distribution),
- $h^2 = .1, .3, .7, .9$  ou 1
- 18 000 ou 5 000 marqueurs sur 3 chr (1 M/chr)
- Bayesian multiple QTL model (Mewissen & Goddard, 2004): effet marqueur et polygénique
- Précision=reliability= $(\text{cor}(\text{GBV}, \text{simulated breeding value}))^2$

# Conclusion : Hayes et al. 2009

- Bonne précision des précision chez la Jersey malgré la petite taille de sa population de référence → Ne faible
- Pour les petites population de référence → population multi espèce = petite amélioration de la précision (13%)
- Pourquoi le gain est-il si faible?  
→ Seuls les SNP en DL très haut sont conservés ou les SNP

# Mat Meth.: Roos et al 2010

- Deux population A et B: 1000 ind /pop
- $h^2 = .1, .3, .7, .9, 1$
- 18 000 ou 5 000 ou 1500 ou 300 marqueurs
- 150 QTL sur 3 chr (1M/chr); distribution exponentielle
- T=6, 30, 300 générations: Au hostein/ Dutch Hostein ou NZ Holstein ou Au Angus
- Methode de Meuwissen et al (2004): tous les marqueurs + effet polygénique
- Reliability=precision= $(r(\text{GBV}, \text{polygénique BV}))^2$

# Resultats: Roos et al 2010

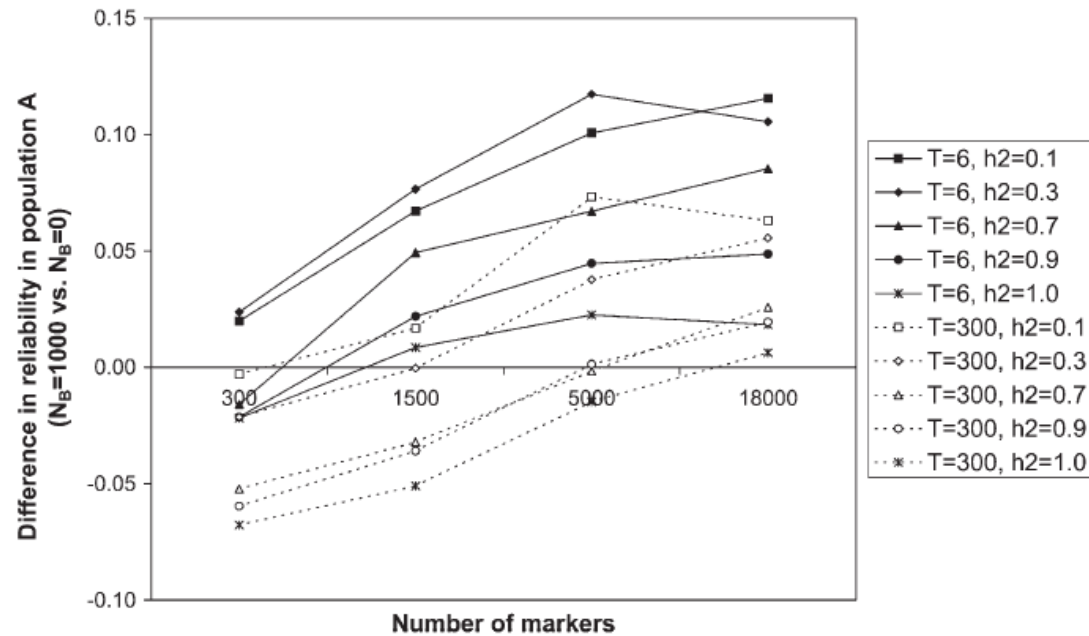


FIGURE 5.—Effect of adding 1000 individuals from population B to the training set ( $N_B = 1000$  vs.  $N_B = 0$ ) on the reliability of genomic breeding values for 1000 individuals from the last generation of population A, in simulations with  $T = 6$  or 300 generations of divergence between populations A and B and different numbers of markers and heritabilities ( $h^2$ ). SEs were  $<0.005$ .

- Différence de précision entre les prédiction de A/A et de A/ A et B:
  - gain  $\uparrow$  avec  $h^2$  et quand divergence faible
  - gain jusqu'à 12%

# Resultats: Roos et al 2010

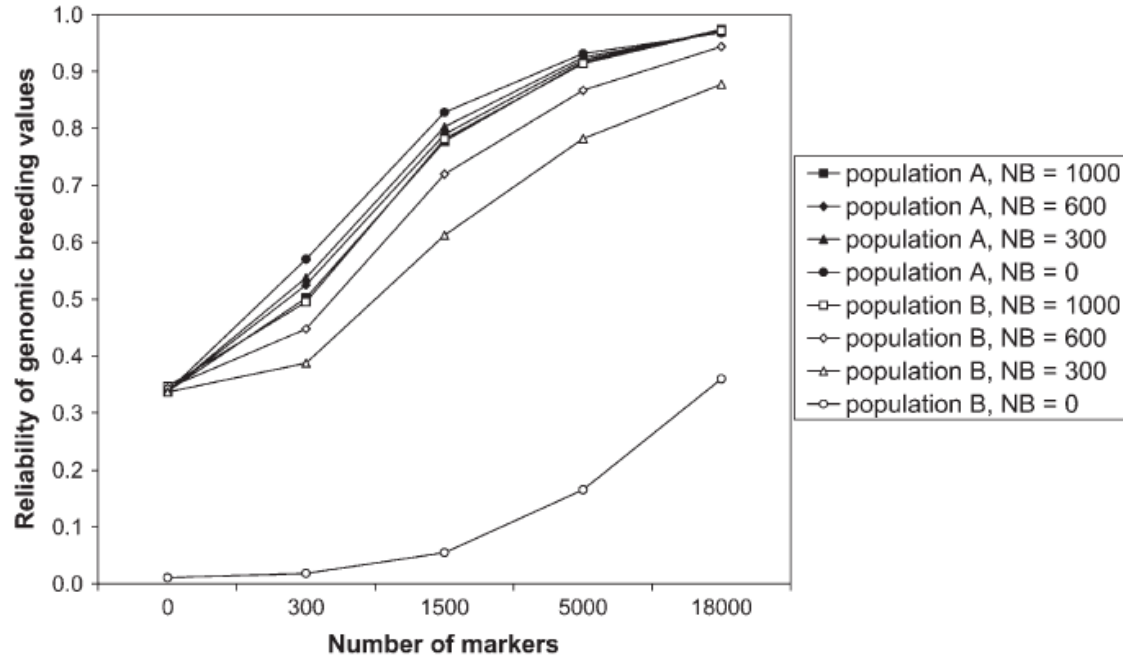


FIGURE 8.—Reliability of genomic breeding values for 1000 individuals from the last generation of populations A and B, after  $T = 300$  generations of divergence between populations A and B, with a heritability of  $h^2 = 1.0$ ,  $N_B = 0, 300, 600,$  or  $1000$  individuals from population B added to the training set, and different numbers of markers. SEs were  $<0.015$  when reliabilities were  $>0.80$  and  $\sim 0.025$  otherwise.

- $h^2=1, T=300$  generat°
  - gain ↑ avec nb de marqueurs 10 à 20%
  - gain ↑ avec prise en compte de la population B dans la population de référence



# Mat Meth: Kizilkaya et al , 2010

- Données génotypes réelles,
- QTL simulés sur certains marqueurs 50 à 500 (effet dans une loi normale)
- $H^2=0,5$
- et phenotypes simulés:
- Plusieurs races: 1086 Angus + 924 (8 races)
- Critère= $\text{cor}(\text{GBV}, \text{EBV simulée})$

# Resultats: Kizilkaya et al , 2010

Table 3. Correlations<sup>1</sup> between true and predicted genotypic values from training in purebred Angus (PB) and validating in individual sire breeds that composed the multibreed (MB) population

QTL scenario <sup>2</sup> and marker panel <sup>3</sup>	Angus	Brahman	Charolais	Hereford	Limousin	Maine-Anjou	Shorthorn	Southdevon
QTL50	239	10	185	78	45	137	97	135
QTL	0.972	0.969	0.956	0.946	0.955	0.967	0.962	0.958
QTL and HLD	0.955	0.919	0.936	0.894	0.908	0.942	0.944	0.929
50K and QTL	0.859	0.858	0.849	0.811	0.810	0.854	0.820	0.836
HLD	0.608	0.339	0.425	0.408	0.412	0.488	0.421	0.481
50K without QTL	0.511	0.227	0.382	0.336	0.355	0.390	0.437	0.403
QTL100								
QTL	0.948	0.918	0.951	0.918	0.938	0.936	0.946	0.930
QTL and HLD	0.919	0.870	0.905	0.847	0.915	0.903	0.904	0.871
50K and QTL	0.704	0.582	0.668	0.548	0.659	0.663	0.657	0.665
HLD	0.623	0.378	0.425	0.426	0.545	0.441	0.422	0.377
50K without QTL	0.386	0.226	0.284	0.158	0.366	0.244	0.363	0.254
QTL250								
QTL	0.882	0.881	0.867	0.816	0.834	0.860	0.825	0.806
QTL and HLD	0.835	0.665	0.781	0.743	0.746	0.805	0.760	0.718
50K and QTL	0.479	0.144	0.415	0.463	0.427	0.415	0.474	0.289
HLD	0.590	-0.197	0.401	0.427	0.242	0.414	0.337	0.310
50K without QTL	0.404	-0.026	0.237	0.300	0.193	0.196	0.306	0.104
QTL500								
QTL	0.819	0.754	0.783	0.739	0.831	0.771	0.808	0.730
QTL and HLD	0.760	0.530	0.704	0.680	0.764	0.705	0.694	0.628
50K and QTL	0.496	0.239	0.380	0.197	0.400	0.260	0.385	0.327
HLD	0.530	0.084	0.331	0.301	0.400	0.365	0.300	0.313
50K without QTL	0.435	0.124	0.262	0.118	0.276	0.174	0.263	0.247

*Pop de référence=1086 Angus*

**La vraie vie = 50k sans QTL → précision de la prédiction dans la même race=40%**

**Mais dans les autres races <0,30**

# Conclusion pour la détection de QTL

- Si carte dense, pas besoin de prendre en compte l'origine racial du DL en SG  
→ extrapolation à la détection de QTL ?
- Modèle BSAM adaptable facilement en détection de QTL facilement (nb paramètres X2) → nécessité de connaître les origines raciales des allèles/haplotypes
- Utilité de connaître DL des races considérées pour utiliser une méthode adaptée