

**Développement, par génétique d'association,  
d'un outil moléculaire de prédiction du poids  
spécifique chez le blé tendre**

**Gilles CHARMET**



**Coordinateur : Olivier Robert - Bioplante**

❖ Identifier, par une nouvelle approche statistique (génétique d'association) des marqueurs moléculaires liés au poids spécifique (PS)

D'autres caractères : le rendement et la floraison ont également été analysés

❖ Mettre au point des équations de prédiction de sélection génomique pour le PS et le rendement et les valider

# Matériel et Méthodes

## Matériel végétal

- ❖ 3 populations de 376 individus
  - ✓ 2 issues de lignées haploïdes doublées (HD1 et HD2)
  - ✓ 1 issue de lignées d'autofécondations (RIL)

## Phénotypage

- ❖ Expérimentations 2009 (HD1) 2010 (RIL) 2011 (HD2)
  - ✓ HD1 : Clermont-Fd (63), Cappelle-en-Pévèle (59), Milly la Forêt (77)
  - ✓ HD2 : Cappelle-en-Pévèle, Milly-la-Forêt, Rennes (35)
  - ✓ RIL : Clermont-Fd, Cappelle-en-Pévèle, Estrées-Mons (80)

## Génotypage

- ❖ Puce à ADN de 5000 marqueurs DArT (Triticarte Pty Ltd)

## Associations génétiques

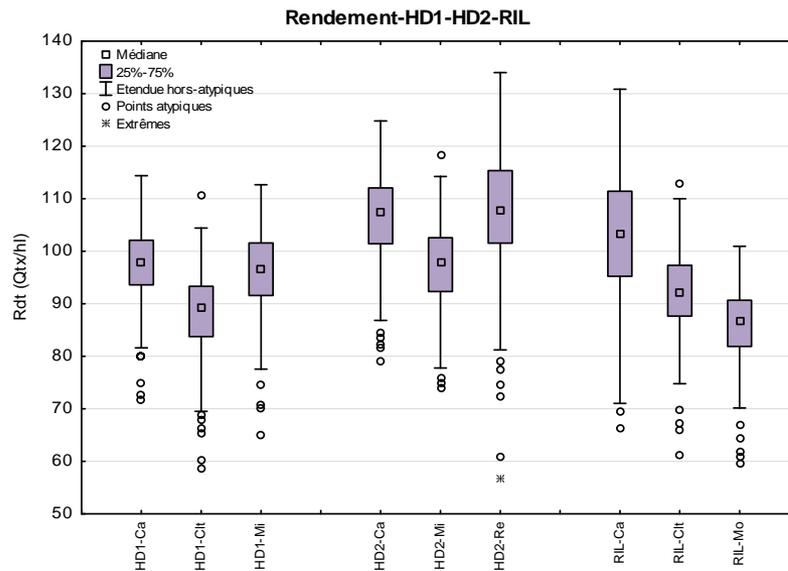
- ❖ Tests des modèles GLM et MLM
  - ✓ Matrice de Structure ou Composantes Principales en co-variable
  - ✓ Matrice des Coefficients de parenté pour modéliser les covariances génétiques en modèles mixtes
  - ✓ Combinaison des deux

## Sélection génomique

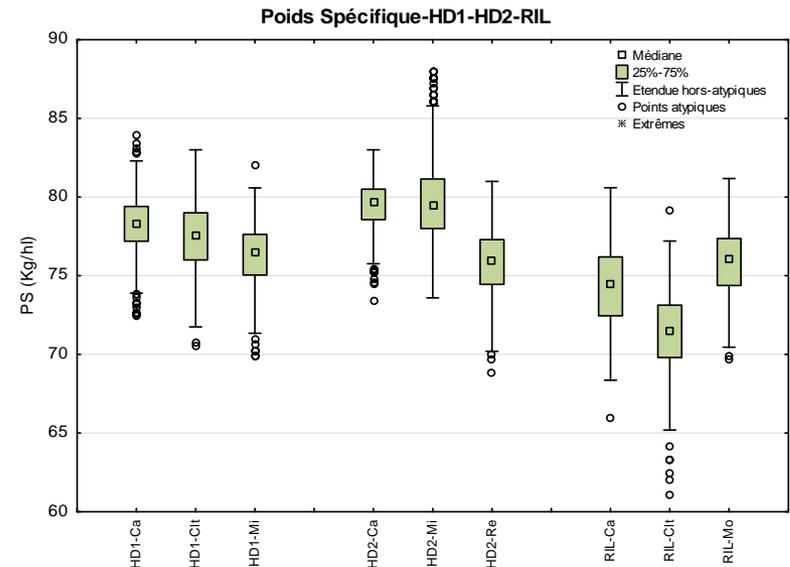
- ❖ Méthodes utilisées
  - ✓ Pedigree-BLUP (G-BLUP) et Ridge Regression-BLUP
    - ✓ Bayesian Ridge Regression et Bayesian LASSO
    - ✓ Random Hilbert Kernel Space et Random forest regression
    - ✓ Validation croisée: 80% « training », 20% « target », 200 runs

# Variabilité génétique

# Variabilité génétique



***Le rendement moyen, relativement fort pour l'ensemble des lieux, indique un matériel végétal de niveau élite. L'étendue, très importante, montre une forte variabilité des 3 populations pour ce caractère***



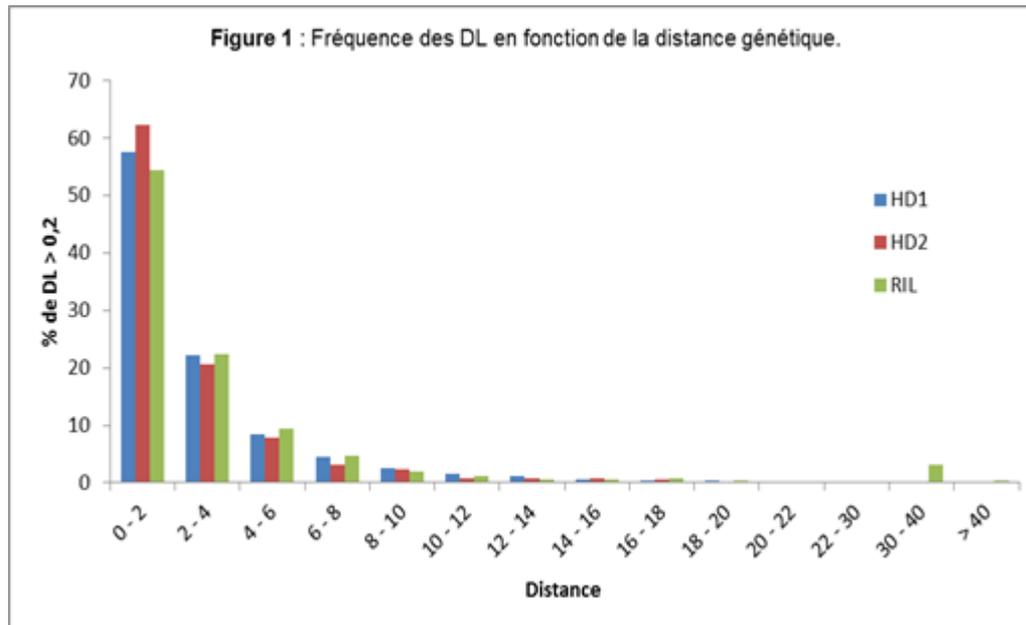
***Le poids spécifique semble plus faible pour la population RIL. Quelle que soit la population considérée, la variabilité est importante, l'étendue égale ou supérieure à 10, va jusqu'à 18 pour la population RIL à Clermont***

# **Marqueurs moléculaires et Déséquilibre de liaison**

# Marqueurs moléculaires et Déséquilibre de Liaison

DARTs polymorphes et communs				
	HD1	HD2	RIL	Communs
	2575	1436	2172	
HD1-HD2	X	X		1435
HD1-RIL	X		X	1951
HD2-RIL		X	X	1235
HD1-HD2-RIL	X	X	X	1154

DL Inter-chromosomique			
Population	Paires	DL $p < 0.001$	DL $p < 0.001$ et $R^2 > 0.2$
HD1	25617	5.16	0.16
HD2	22347	5.00	0.13
RIL	24043	7.35	0.10

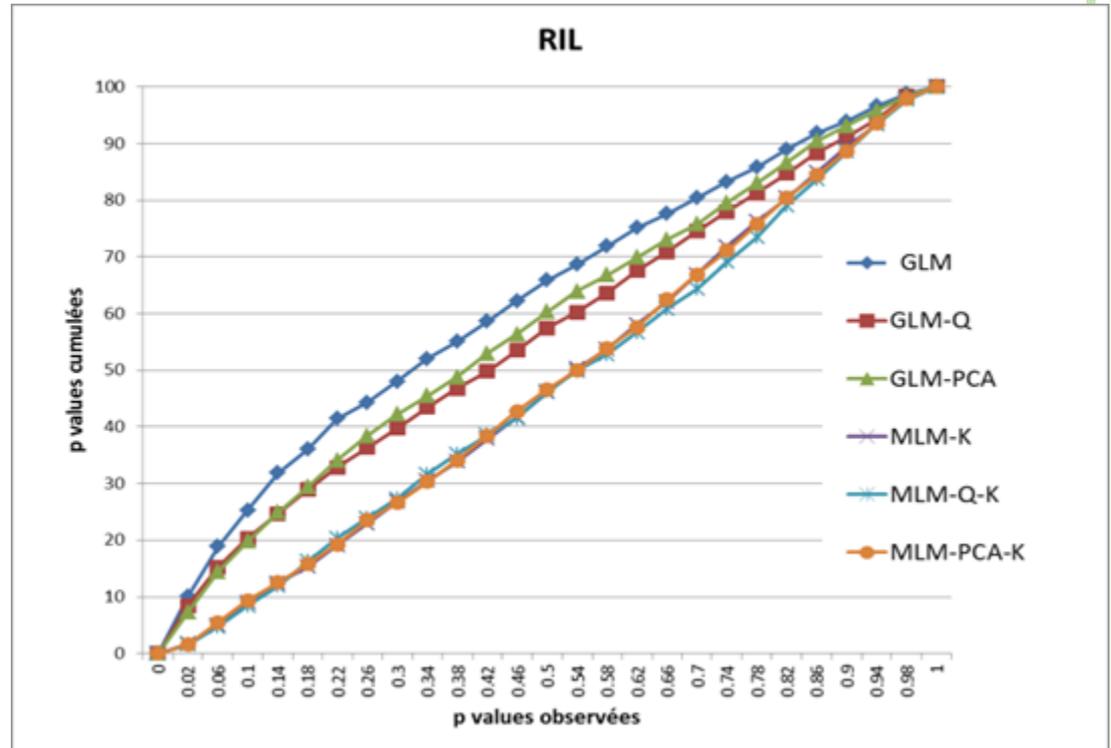


- ✓ **DL inter : RIL > HD pour  $p < 0.001$  > opposé à ce qui est attendu entre Haplo-diploïdisation et Auto fécondations**
- ✓ **avec  $p < 0.001$  > mais un  $R^2 > 0.2$  peu de DL => absence de structure**
- ✓ **DL intra : comportement semblable pour les 3 populations**
- ✓ **Présence d'une empreinte de sélection à 30 – 40 cM ?**

# **Associations génétiques**

# Associations génétiques

Recherche du modèle le mieux adapté à partir de la floraison : caractère à hérabilité forte (0,78 - 0,92) et donc peu influencée par l'interaction génotype x milieu



# Associations génétiques

Régions très impliquées dans le déterminisme du poids spécifique (PS) et du Rendement (RDT), identifiées à partir des 3 populations

1D		PS
wPt-3738	32.07	HD1-HD2-RIL
wPt-1387	39.51	
wPt-7946	40.50	
wPt-741323	42.59	
wPt-664824	44.48	
wPt-0487	44.50	
wPt-671990	44.89	

4A		PS
wPt-744256	71.37	HD1-HD2-RIL
wPt-5172	71.69	
wPt-669203	71.96	
wPt-5374	72.08	
wPt-4828	72.15	
wPt-9883	72.24	
rPt-0238	72.34	

3A		PS
wPt-9422	81.83	HD1-HD2-RIL
wPt-4725	82.08	
wPt-741749	84.00	
wPt-730156	94.17	
wPt-4128	94.91	
tPt-9901	95.37	
wPt-9761	97.37	
wPt-743015	106.31	
wPt-1888	106.82	

3B		PS
wPt-1191	46.60	HD1-HD2-RIL
wPt-742225	46.75	
wPt-742060	46.80	
wPt-743395	48.80	
wPt-8915	52.27	
wPt-11295	58.58	
wPt-5939	58.95	

3D		PS
wPt-667560	36.72	HD1-HD2-RIL
wPt-731416	36.75	
wPt-744794	42.82	
wPt-732381	42.84	
wPt-7535	42.97	
wPt-9662	50.85	

6B		PS
tPt-3506	25.54	HD1-HD2-RIL
wPt-1325	31.25	
wPt-8721	33.16	
wPt-2810	34.80	
wPt-9952	35.96	
wPt-1264	36.03	

6A		PS	RDT
wPt-9858	0.00	HD1-HD2-RIL	HD1-HD2-RIL
wPt-666927	2.64		
wPt-669271	3.35		
wPt-1742	3.57		
wPt-9832	3.99		
wPt-667740	8.62		
wPt-6904	8.99		
wPt-8266	16.35		
wPt-0168	18.20		
wPt-1060	19.20		
wPt-730460	20.06	HD1-HD2-RIL	HD1-HD2-RIL
wPt-9132	20.34		
wPt-733151	25.28		
wPt-729839	38.81		
wPt-2206	87.23		
wPt-4229	91.41		
wPt-5696	91.47		
wPt-666773	92.17		
wPt-731786	92.48		
rPt-16189	94.77		
wPt-731524	96.00		

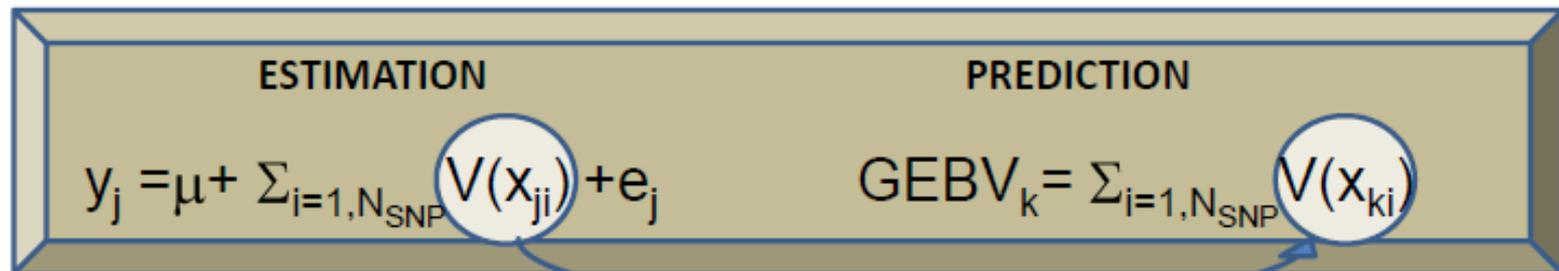
1B		RDT
wPt-0308	13.98	HD1-HD2-RIL
wPt-1637	24.70	
wPt-4325	25.05	
wPt-5313	28.63	
wPt-9973	30.20	
wPt-4655	32.70	
wPt-3451	35.92	

7B		RDT
wPt-7887	116.93	HD1-HD2-RIL
wPt-5975	117.00	
wPt-8615	120.00	
wPt-9992	123.00	
wPt-8989	128.00	
wPt-3939	135.05	
wPt-665293	139.26	
wPt-1066	142.36	

# **Prédiction (ou sélection) génomique**

# La SG : Comment ça marche ?

- **Population de référence : ESTIMATION**
  - 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 : PREDICTION**
  - Population présentant les mêmes associations
  - D'individus génotypés
  - Prédiction de la valeur à partir des effets estimés



# Avantages de la SG sur la sélection phénotypique

h ou précision du prédictueur

$$DG = i h s_G / L$$

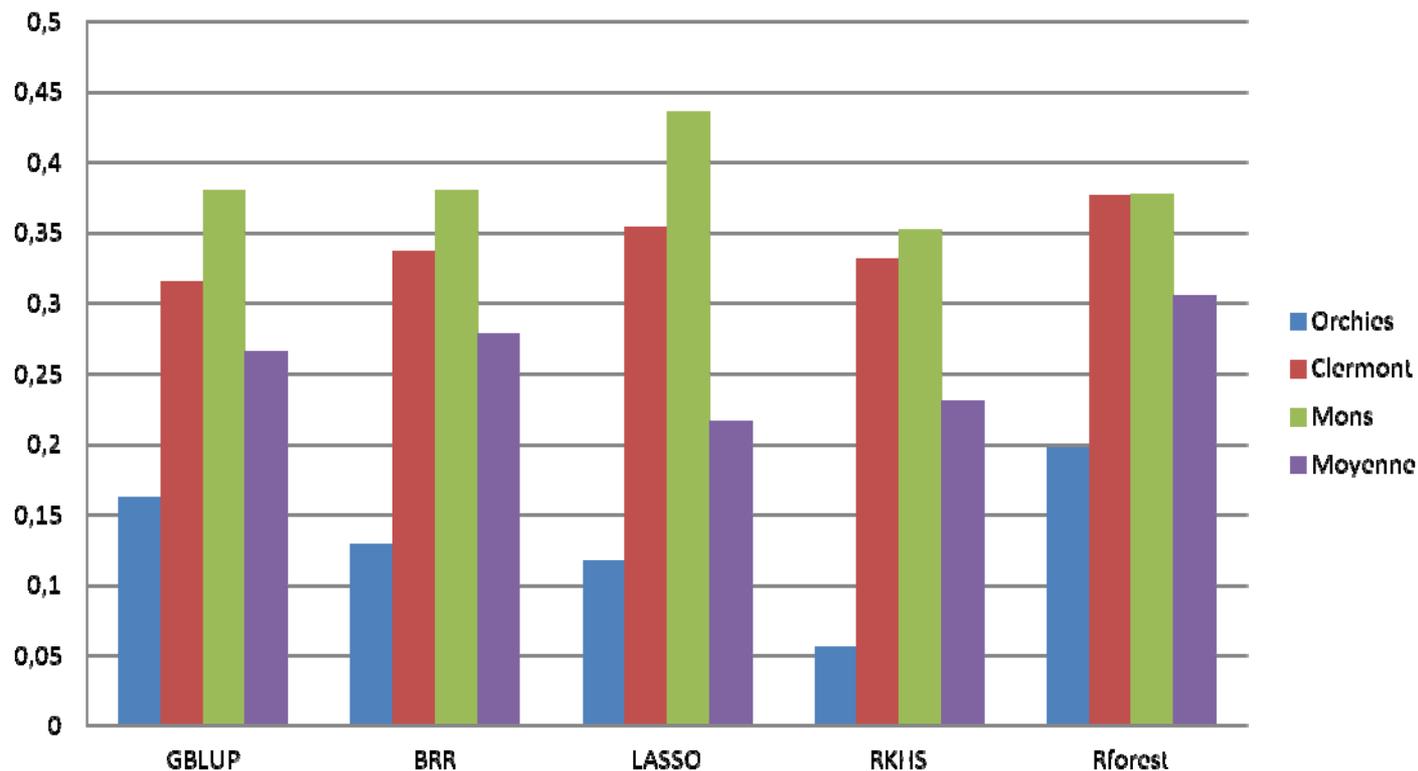
Intensité de sélection.  
Peut être augmentée si  
Coût génotypage  $\ll$  phénotypage

Longueur du cycle.  
Peux être raccourci si  
Sélection précoce

# Prédiction génomique du rendement (RILs): validation croisée

Précision des prédictions comparable aux données de la littérature, mais très variables entre lieux, plus qu'entre les méthodes

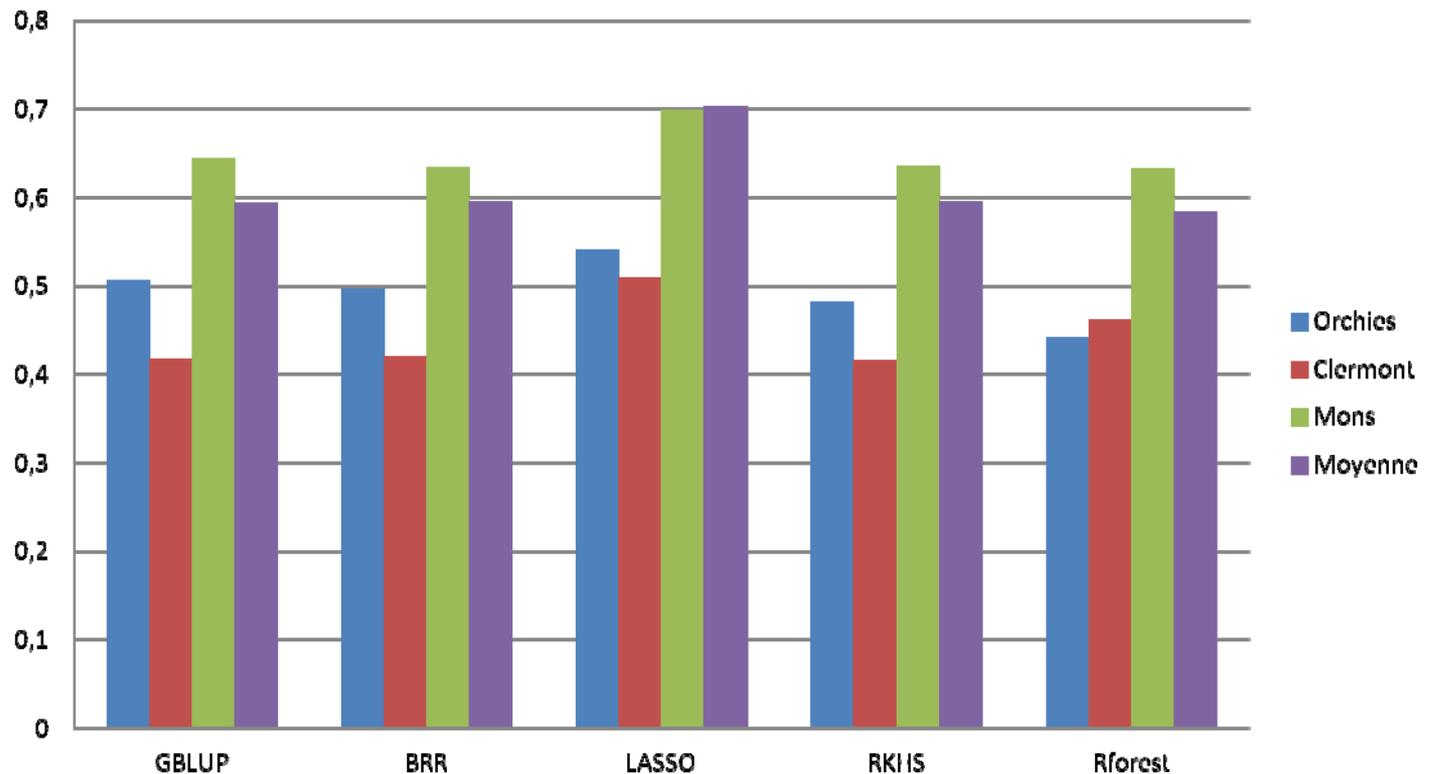
**Rendement: cor(GEBV, pheno)**



# Prédiction génomique du poids spécifique (RILs) : validation croisée

Précision des prédictions plus homogène et plus intéressante: utilisable pour sélection sur plantules?

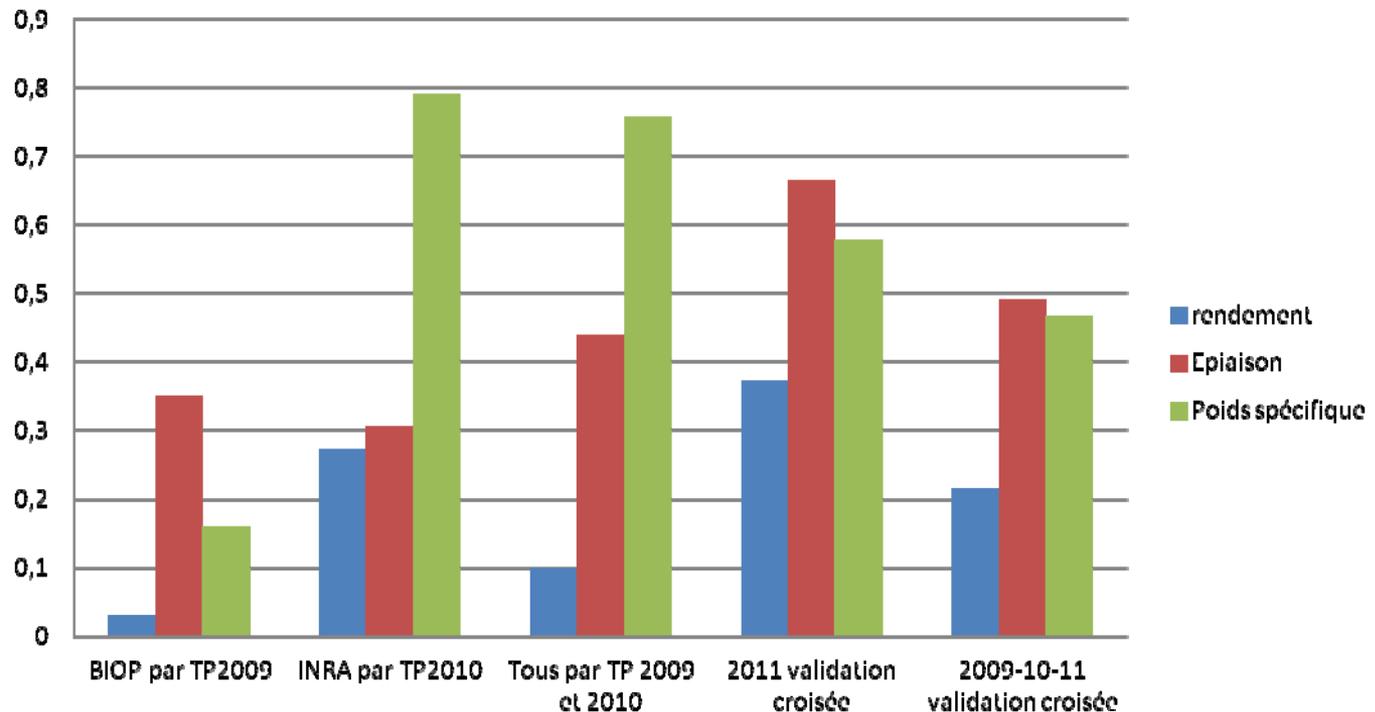
**PS: cor(GEBV, Pheno)**



# Prédiction génomique du PS: HD2 par différents training panels

Prédictions entre panels aussi précises (parfois) plus que les prédictions « intra » (VC)

**Cor(GEBV, pheno) sur panel 2011 selon training panel (TP)**



# Conclusion

- Etendue du DL intra et peu de DL à longue distance : les 3 panels étaient de bons outils pour la GWAS
- Utilisation conjointe de structure/CP et des modèles mixtes génétiques : bon contrôle des faux positifs
- 8 QTL communs aux 3 panels pour le PS et 5 pour le rendement
- Prédiction génomique du PS relativement précise et stable selon le panel utilisé comme population de référence
- La prédiction génomique pour le rendement doit encore être améliorée (Breedwheat...)

# Remerciements

- Bioplante

- Olivier Robert,
- Ellen Goudemand,
- Laure Duchalais,
- Valérie Laurent,
- Denis Béghin

- INRA

- Jacques Bordes,
- Annie Lapierre,
- Laëtitia Chevarin,
- Marie-Reine Perretant,
- François-Xavier Oury,
- Bernard Rolland,
- Laurence Moreau,
- Alain Charcosset
- Eric Storlie,
- Gilles Charmet