

**Journées Scientifiques du Groupe Céréales
Montpellier 3 et 4 Avril 2013**



Couplage entre phénotypage, modélisation et génétique

Situation existante et perspectives (et limites)



VINCENT ALLARD
INRA, UMR 1095 GDEC
vincent.allard@clermont.inra.fr

03/04/2013

Sources

Bénédicte Quilot-Thurion : “*Intégration du contrôle génétique dans les modèles*”. Présentation faite dans le cadre de l'école chercheurs “*Conception d'idéotypes de plantes pour une agriculture durable*” 22-25 octobre 2012, Domaine de Seillac.

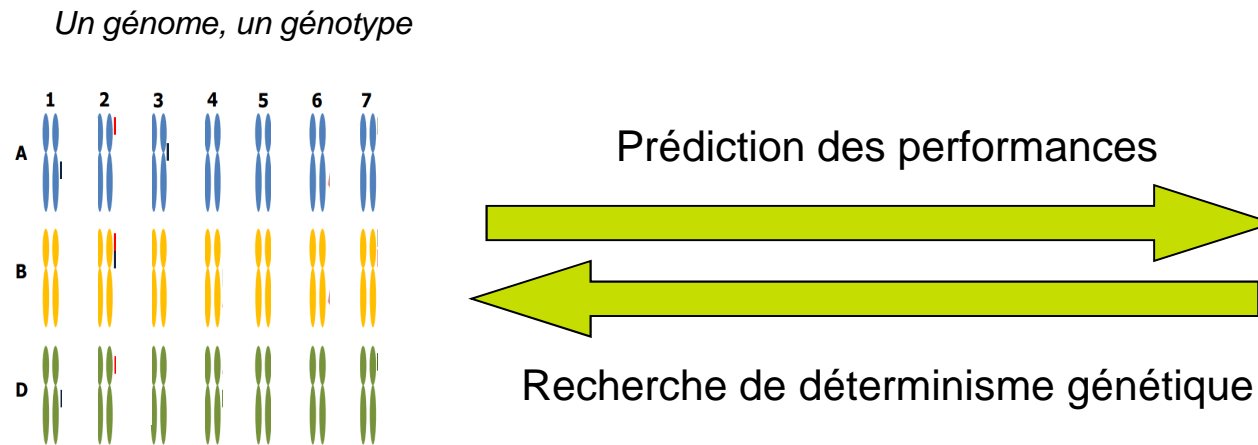
Matthieu Bogard : Post-doctorant INRA UMR GDEC / CSIRO, projet KBBE Adaptawheat
Modélisation basée sur les gènes de la précocité du blé tendre.

_01

Introduction

Les interactions GxE, c'est grave mais ca se soigne (un peu)

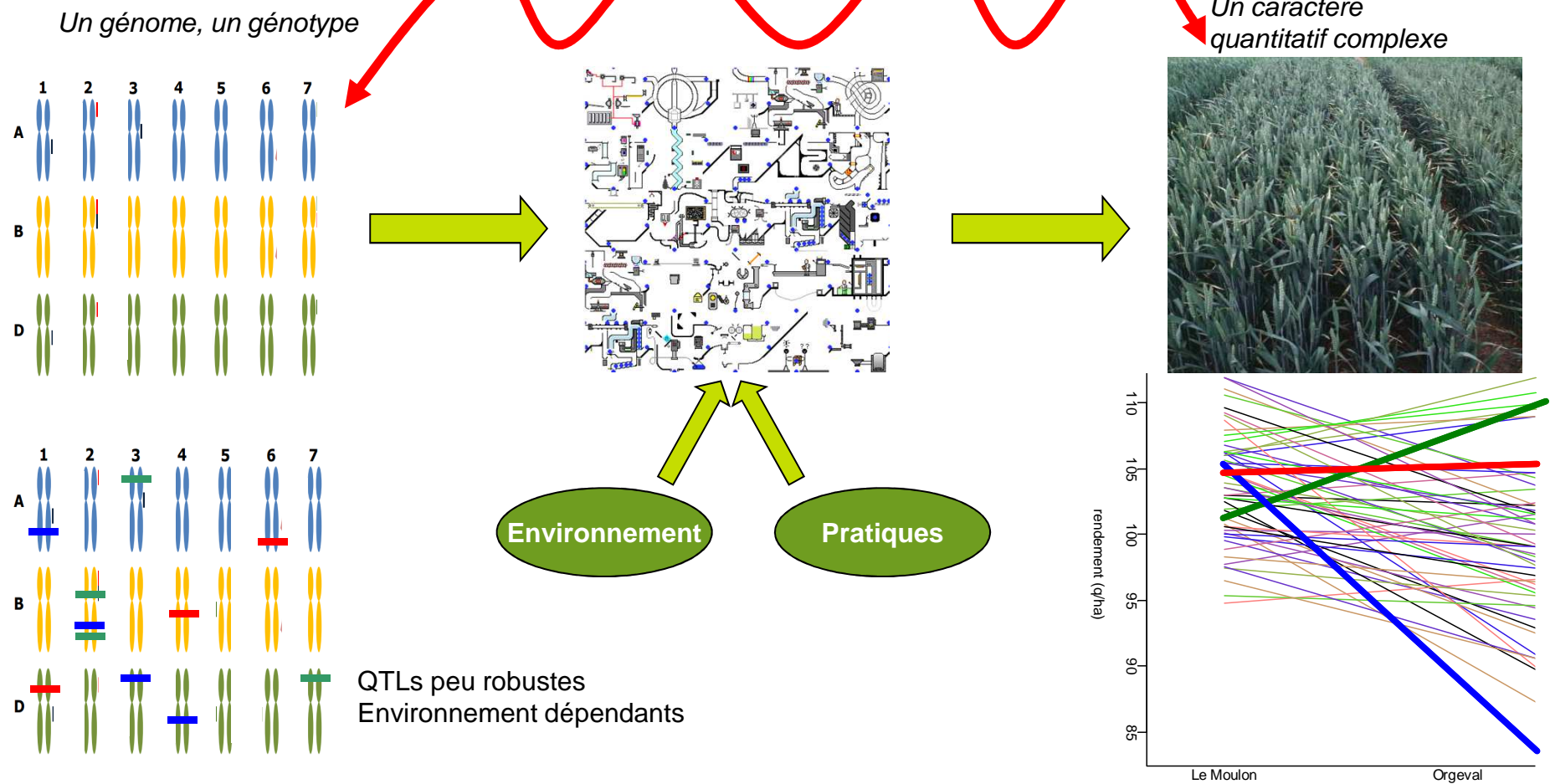
Les interactions GxE



*Un caractère
quantitatif complexe*

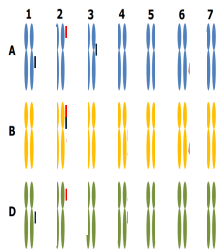


Les interactions GxE



Modéliser les interactions GxE

Un génome, un génotype

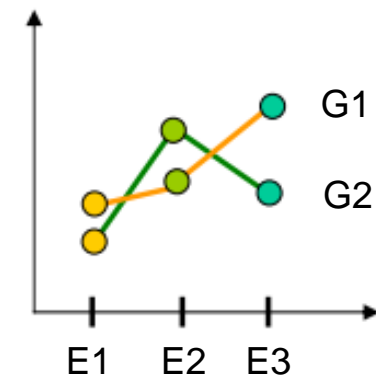
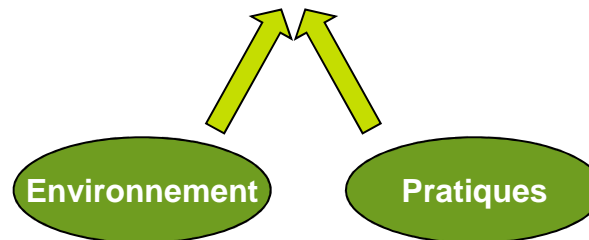


**Paramètres
génétiques**

Modèle écophysologique



*Un caractère
quantitatif complexe*

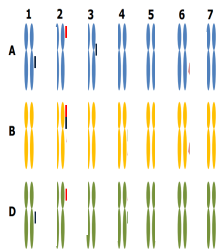


Modéliser les interactions GxE : Pourquoi

- Etudier les interactions Génotype x Environnement x Management
- Simuler le phénotype d'un génotype virtuel avec une combinaison d'allèles donnée dans n'importe quel scénario cultural et climatique
- Rechercher du déterminisme génétique robuste (indépendant de l'environnement)
- Vers la sélection assistée par modèle ou 'virtual breeding'

Modéliser les interactions GxE

Un génome, un génotype

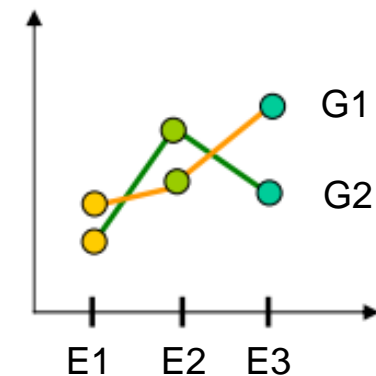
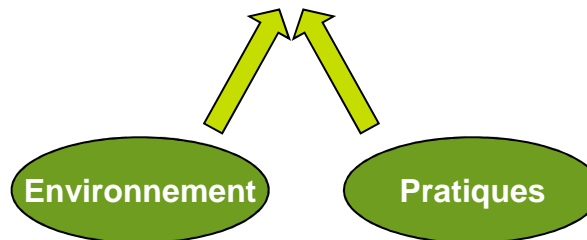


**Paramètres
génétiques**

Modèle écophysologique



*Un caractère
quantitatif complexe*



Décrire les génotypes: le problème de la paramétrisation

Deux voies de paramétrisation des modèles

Estimation directe des paramètres (**Paramètre intégratif**) cf. analyse de sensibilité ou démarche amont poussées d'identification de processus clés

Exemples:

- *Reymond et al. 2003; 2004: croissance de la feuille de maïs sous contraintes hydrique et thermique*
- *Yin et al. 2005: floraison de l'orge*

Optimisation (**Paramètre caché**)

Exemples:

- *Letort et al. 2008: modèle GreenLab*
- *Luquet et al. 2012: vigueur initiale du riz*
- *Bogard et al. Unpublished exemple à suivre*

_02

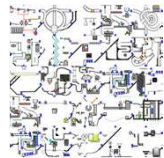
Exemple

Modélisation basée sur les gènes de la précocité du blé tendre

Stratégie utilisée

Modèle de phénologie Arcwheat

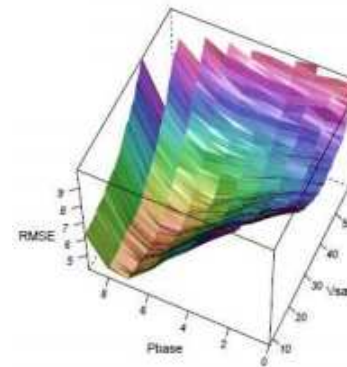
1 panel d'association



$\left\{ \begin{array}{l} P_{base} \\ V_{sat} \end{array} \right\}$ 2 paramètres génétiques

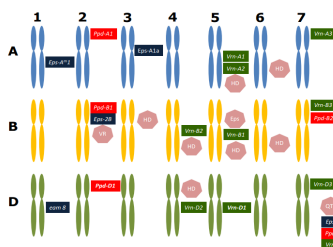
Core collection

1 Optimisation des paramètres



Utilisation des données observées

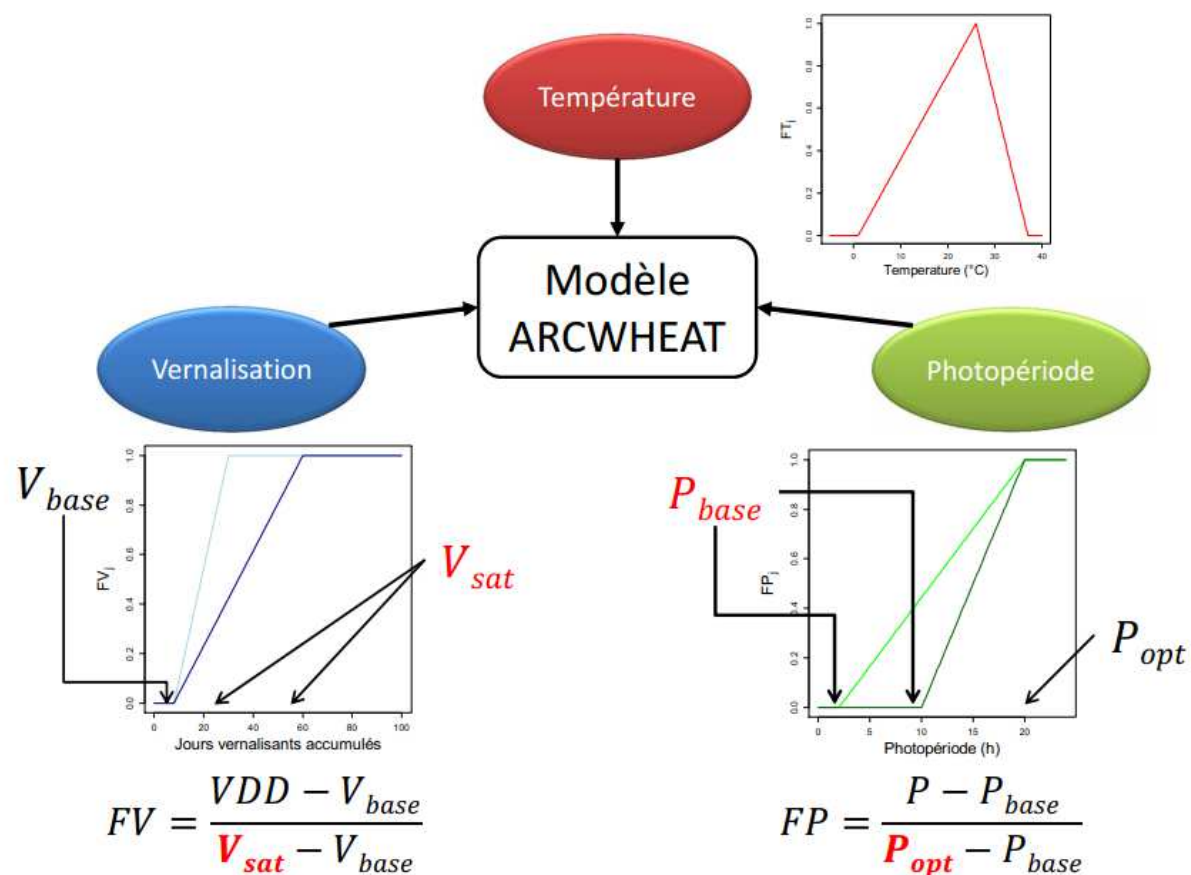
2 Détection de QTL ROBUSTES



3 Prédiction des paramètres à l'aide des QTLs

$$\begin{aligned} P_{base} &= \alpha M k_1 + \alpha M k_2 + \dots + \varepsilon \\ V_{sat} &= \alpha M k_1 + \alpha M k_2 + \dots + \varepsilon \end{aligned}$$

Modèle utilisé



Données phénotypiques

Core collection (234 accessions)		
Lieu	Date de semis	
Clermont-Ferrand	27/10/2004	
	08/11/2005	
	20/11/2008	
	20/02/2009	
	20/03/2009	
	22/04/2009	
Le Moulon	20/10/2003	
	04/04/2005	
	13/02/2004	
	12/03/2004	
	05/04/2004	
	20/10/2004	
	10/03/2005	
	20/10/2005	
	17/02/2006	
	20/03/2006	
	07/04/2006	

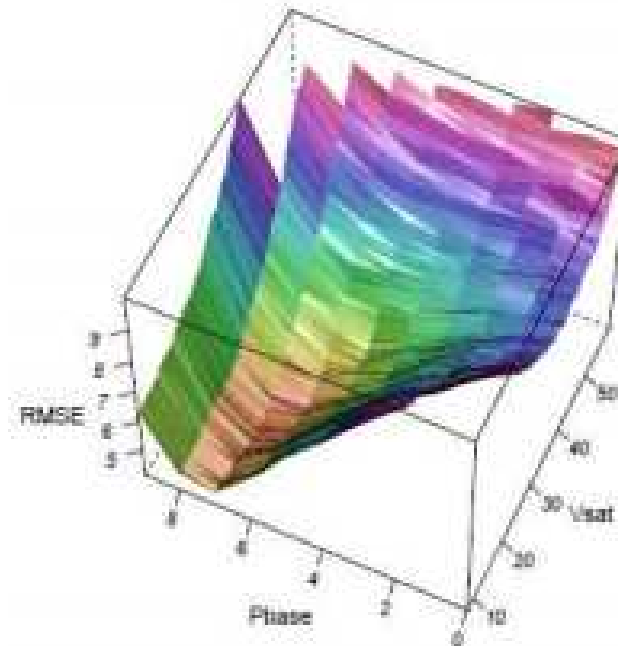
Unvernalized spring sowing

4 week vernalization

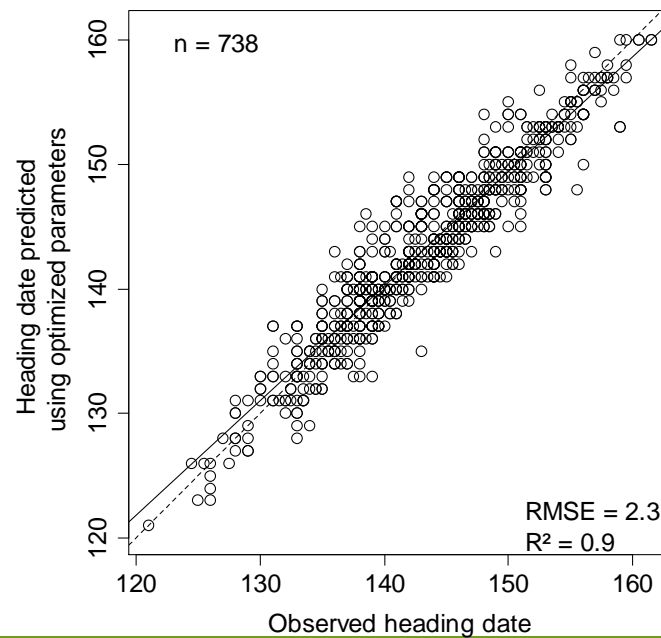
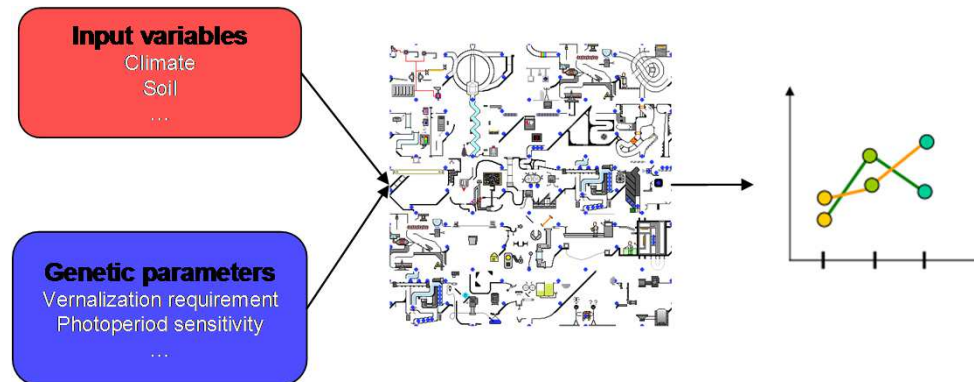
8 week vernalization

Optimisation des paramètres

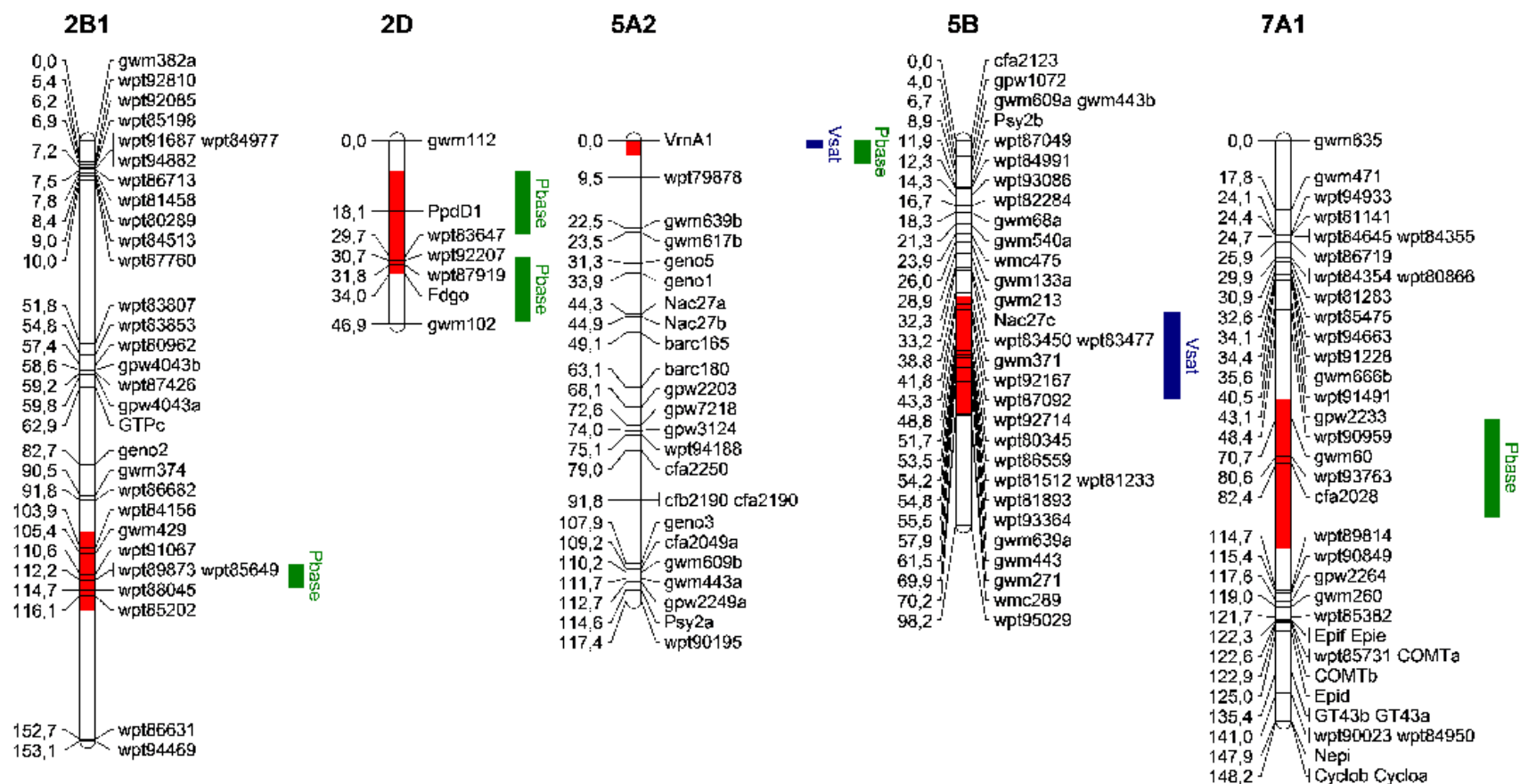
- Objectif: Trouver le meilleur binôme de paramètres en terme de pouvoir prédictif dans l'ensemble des environnements
- Exploration de l'espace
 - Pbase dans [0;12]
 - Vsat dans [0; 120]
- Calculs de RMSE pour les 1573 binômes



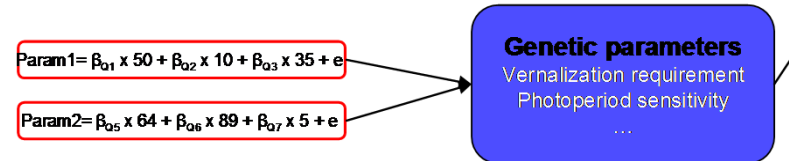
Capacité prédictive du modèle à partir des paramètres optimisés



Génétique d'association sur les paramètres estimés

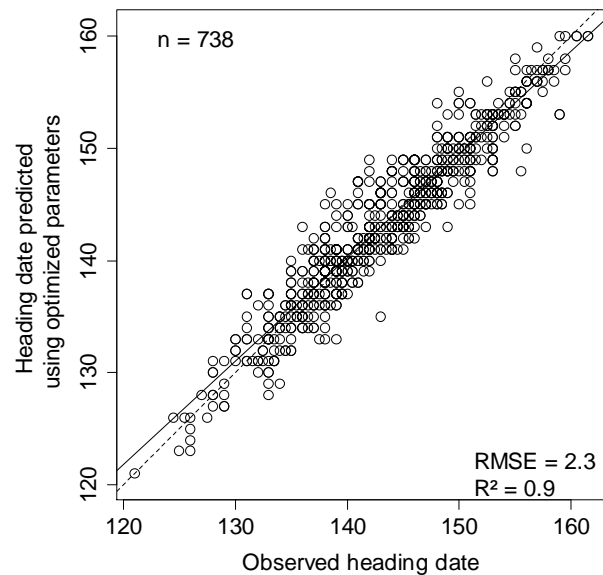
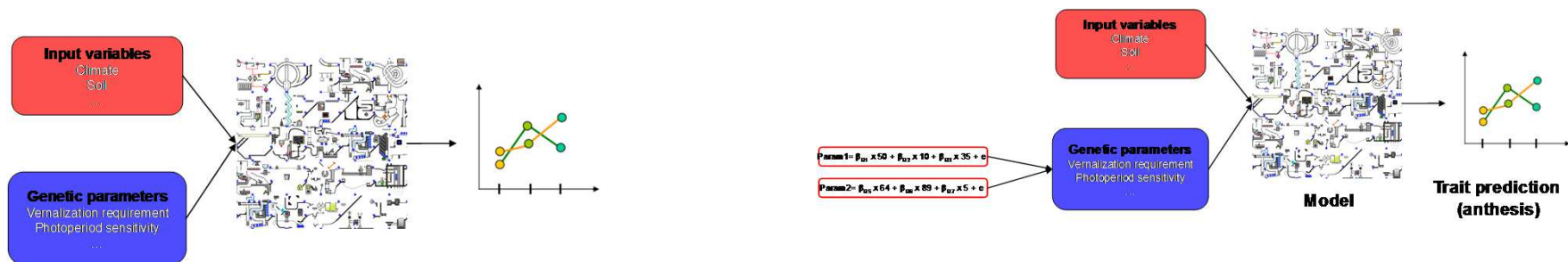


Prédiction des paramètres par régression multiple

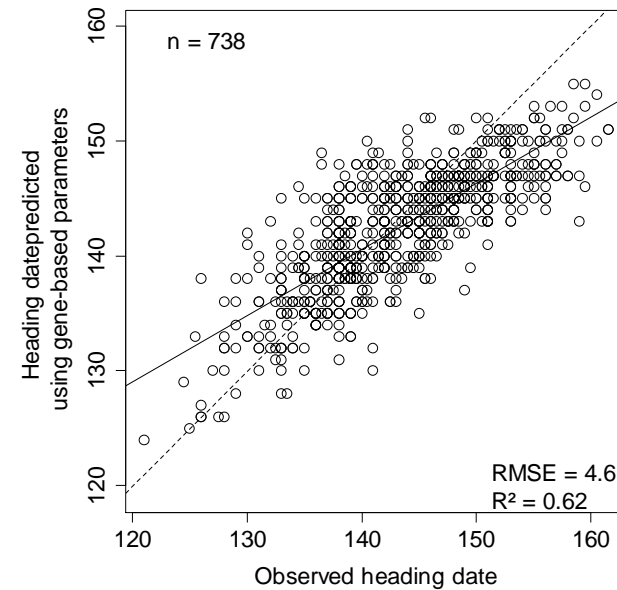


Parameter	Marker	Chrom.	Allele	Coeff.	Std error	p-value
P_{base}	Intercept	-	-	5.6	0.33	<0.001
	PpdD1.PromDel (1)	2D	2	-1.7	0.21	<0.001
	EPI.1B.Y.433 (C)	1B	T	0.5	0.27	0.05
	wPt4515 (0)	7A	1	0.6	0.28	<0.05
	wPt9380 (0)	1D	1	-0.5	0.18	<0.01
	GDH.3.R.1656 (A)	3	G	-0.8	0.36	<0.05
	gluA1.1.Y.1667 (C)	1A	T	0.7	0.18	<0.001
	Spa2prom.5D.Y.287 (C)	5D	T	-0.7	0.30	<0.05
	PHD20.4B.Y.544 (C)	4B	T	-0.6	0.19	<0.01
V_{sat}	Intercept	-	-	40.7	8.3	<0.001
	vern.5B.Sins.8761 (del)	5B	ins	28.2	4.8	<0.001
	Vrn.A1pr (11)	5A	22	27.6	6.0	<0.001
			55	0.6	8.4	0.94
	FdGogat.2D.Y.545 (C)	2D	T	-19.1	4.6	<0.001
	gpw4129 (339)	7D	341	-6.8	5.0	0.18
	GT43.7D.M.322 (A)	7D	C	11.5	4.8	<0.05
	wPt1370 (0)	5A	1	-11.9	4.2	<0.01

Capacité prédictive du modèle basé sur les QTL



RMSE x 2



Validation sur génotypes indépendants

103 génotypes de la Core collection non utilisés dans la calibration

Location	Sowing date	n	RMSE	b	r ²
Clermont	27/10/2004	103	3.7	1.00	0.61
	08/11/2005	49	5.2	1.00	0.39
Joze	29/10/2006	63	7.3	1.04	0.48
	25/10/2007	63	10.1	1.06	0.46
Le Moulon	26/10/2006	63	5.5	1.00	0.56
	23/10/2007	63	6.2	1.03	0.54
Mons	17/10/2006	63	7.4	1.03	0.53
	22/10/2007	63	8.2	1.05	0.51



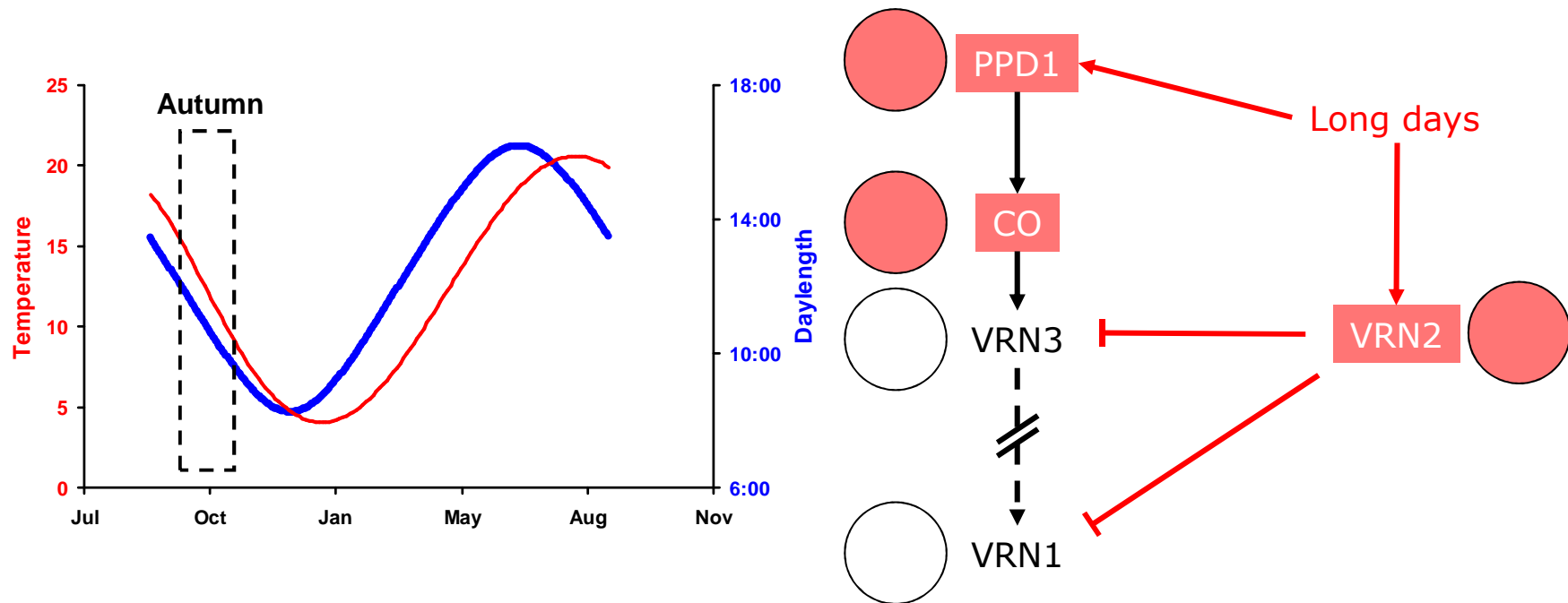
Explosion de la RMSE. Problèmes d'haplotypes non représentés?
Effets épistatiques du fond génétique non détectés?

03

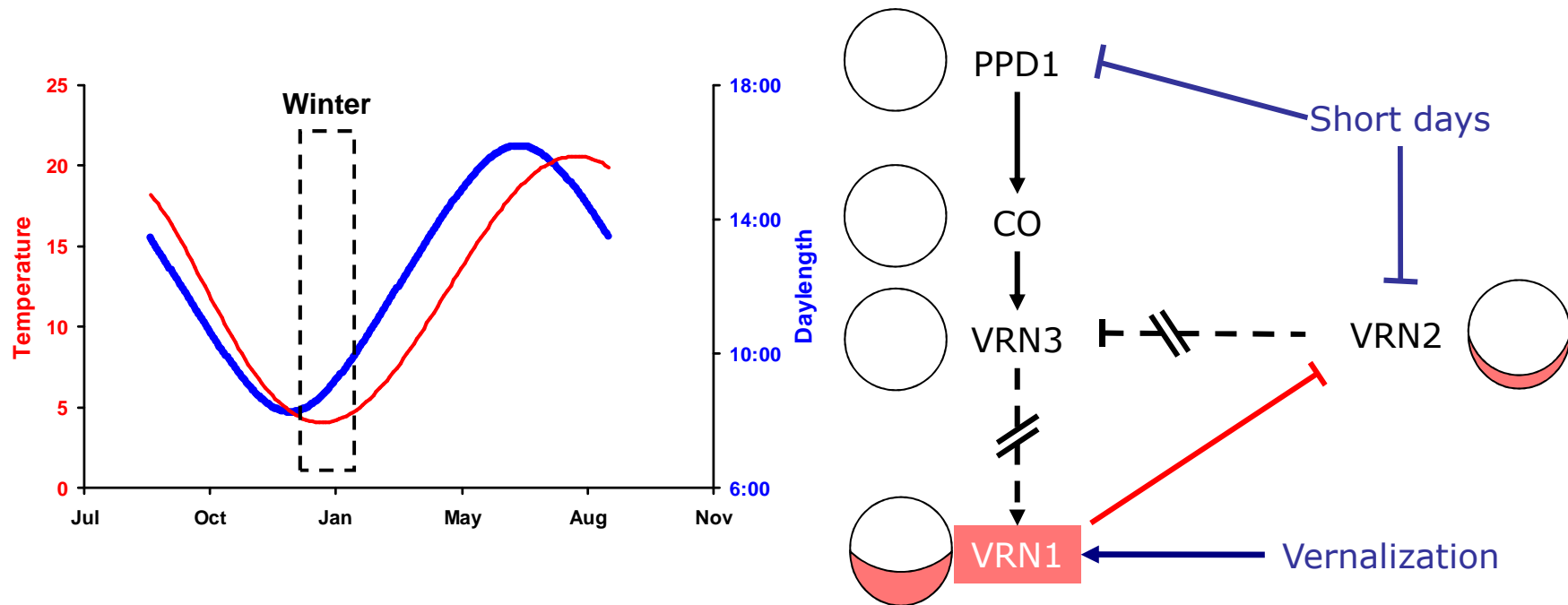
Vers un gene based model au sens strict

Une fuite en avant vers la complexité?

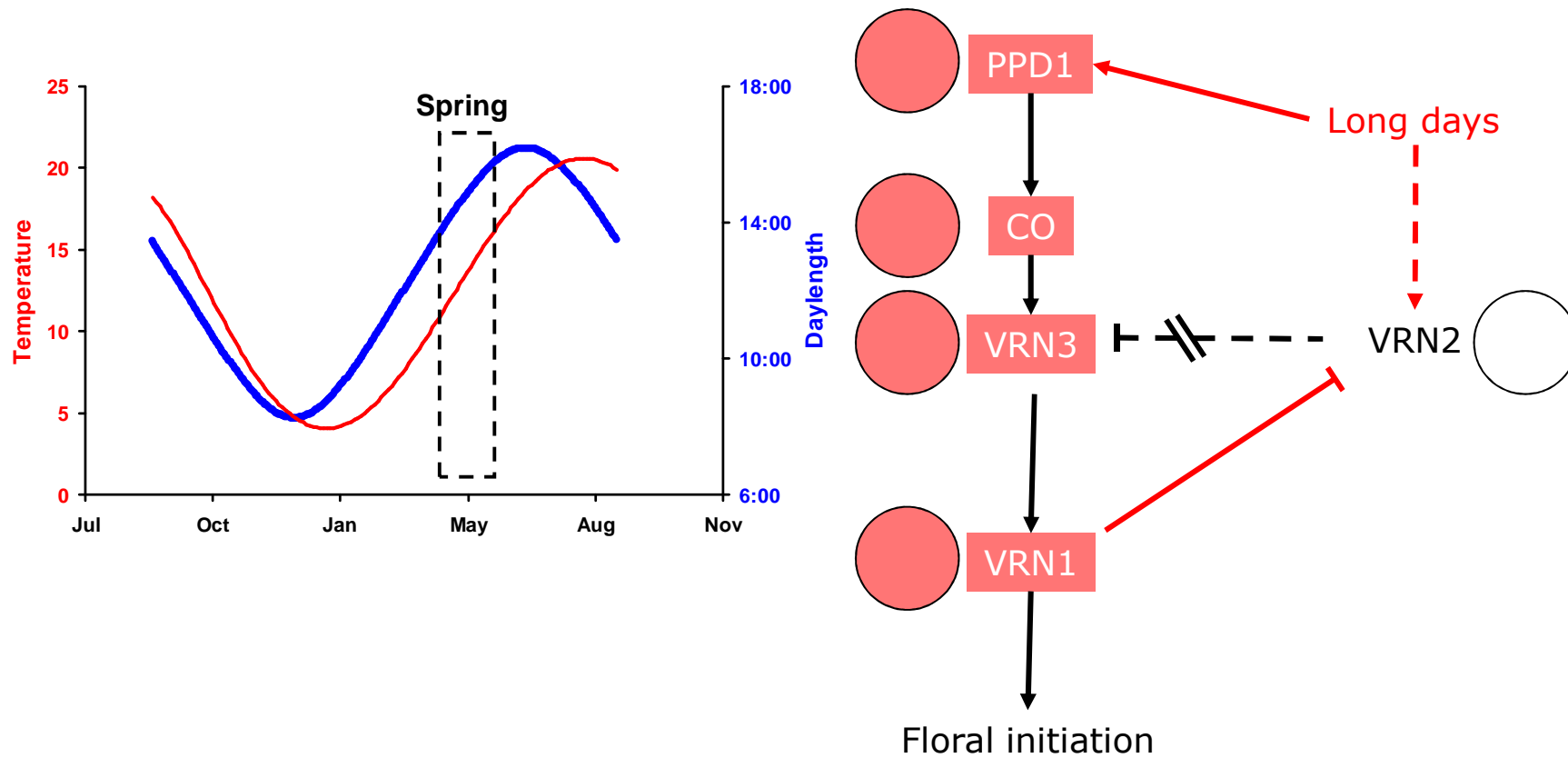
Un réseau de gènes bien connu



Un réseau de gènes bien connu

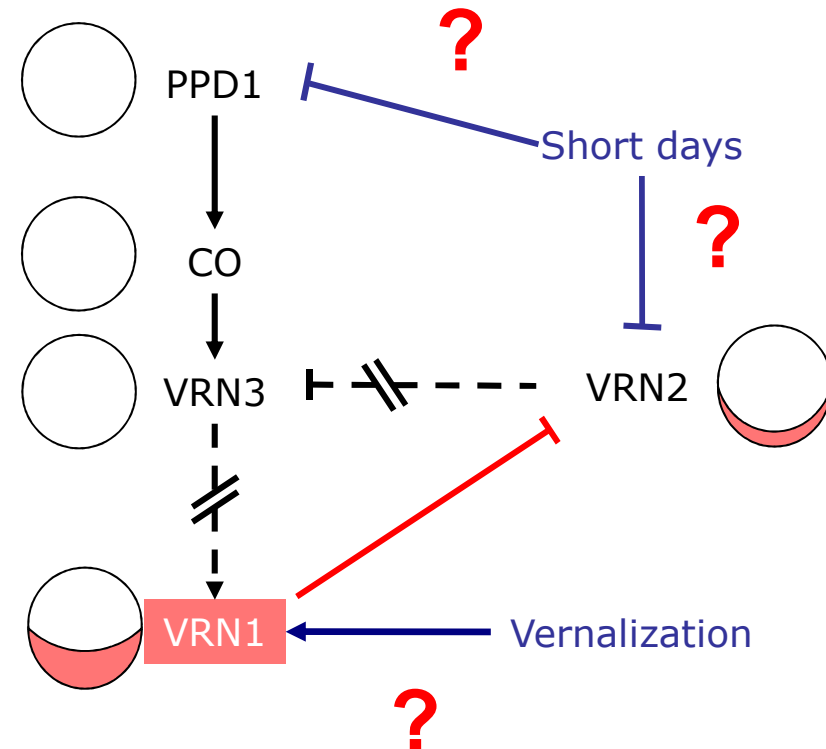


Un réseau de gènes bien connu

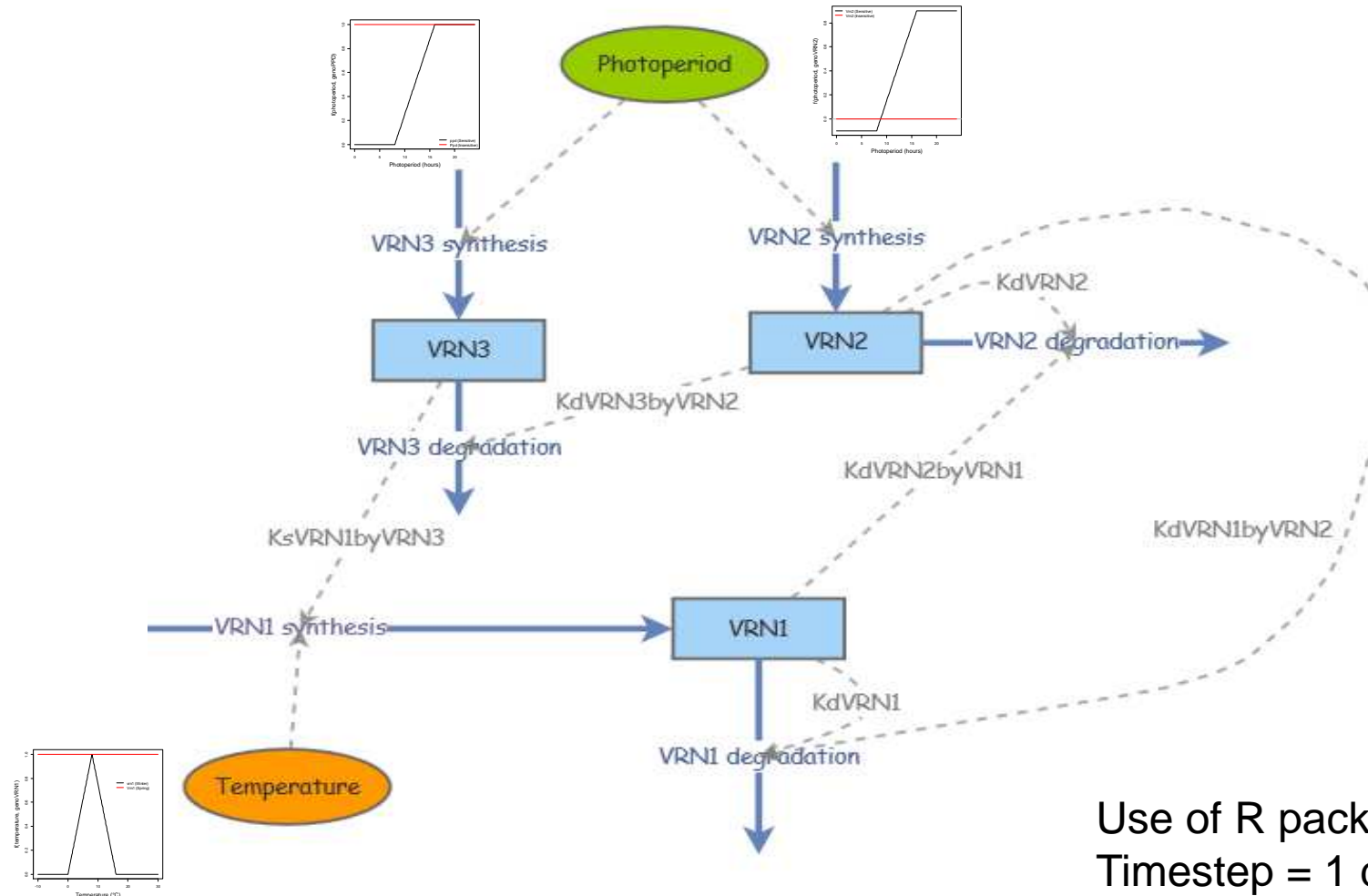


Mais sans information quantitative

- Introduire des courbes de réponses de l'expression aux facteurs de l'environnement
- L'expression individuelle a des propriétés émergentes à l'échelle du réseau
- Les effets alléliques peuvent être décrit de manière explicite

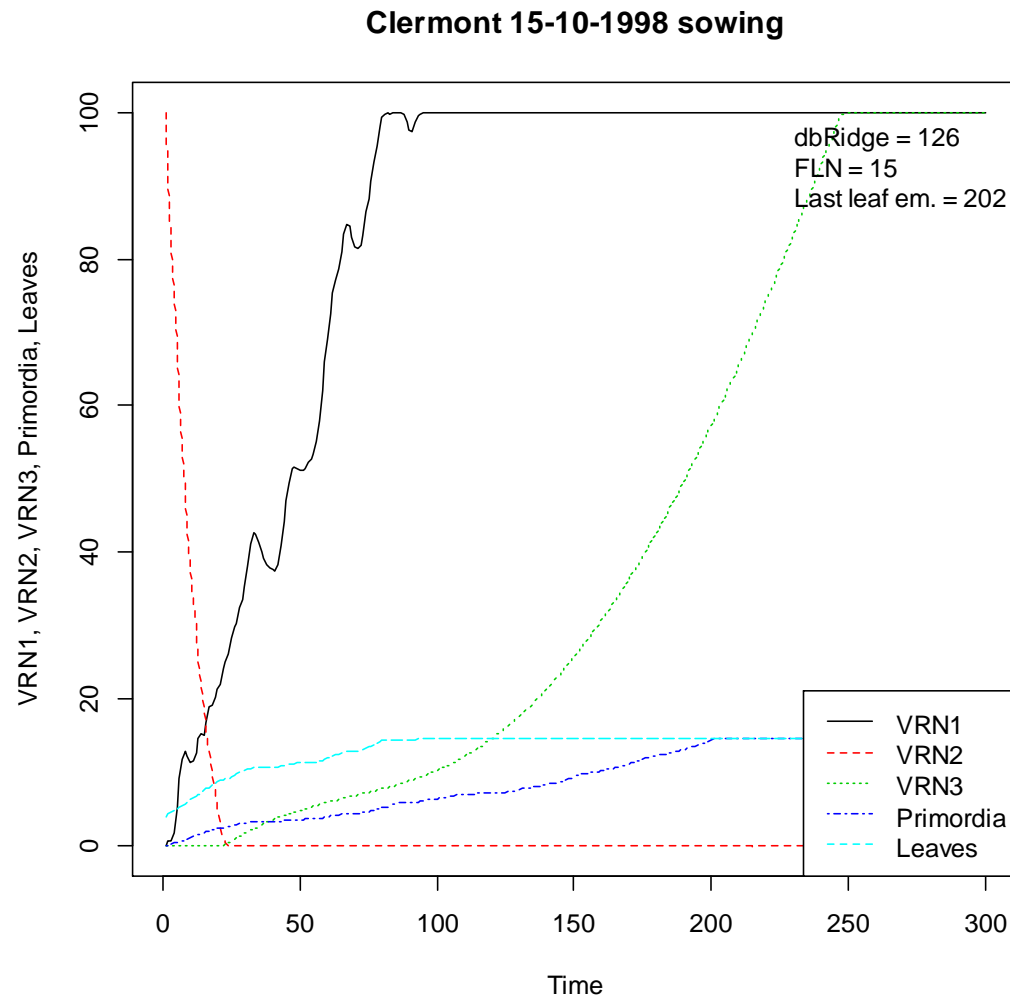


Comment modéliser dynamiquement un réseau



Use of R package Simecol
Timestep = 1 day

Propriétés émergentes au niveau du réseau



_03

Perspectives et limites

Qu'avons nous fait exactement?

Un modèle prédictif de la précocité basé sur des gènes (QTL) (modèle sur un caractère **complexe...mais simple**)

- Capacité prédictive (5-10 jours) : intérêt pour la modélisation de type “land use”? Pour l'idéotypage? Pour la sélection?
- Détection de QTLs “robustes”: colocalisation avec des QTL connus. Question de l'usage de ces QTLs et de leur introduction dans des schémas de sélection.
- Intérêt propre de la formalisation des connaissances

Quelles perspectives?

Nécessité d'une réflexion pluridisciplinaire sur les attendus de ce type d'approche en terme de

- Précision
- De mode de représentation de la variabilité génétique

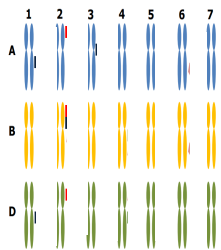
⇒ Un modèle, une démarche, un usage.

Devenir collectivement réalistes quant à l'impact possible de telles démarches

Réaliser de réelles estimations « cout / bénéfice » pour les usages finalisés.

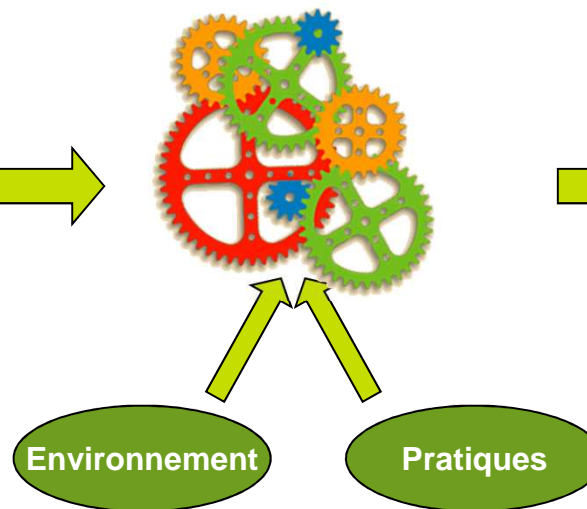
Quelles perspectives?

Un génome, un génotype



**Paramètres
génétiques**

Modèle écophysologique



*Un caractère
quantitatif complexe*



Complexité du déterminisme
Problème des panels
d'apprentissage vs. validation

Introduire suffisamment de
complexité pour être en situation de
générer les IGE utiles mais rester
suffisamment simple pour limiter
les problèmes de paramétrization
(SG sur l'ensemble des
paramètres?)

Peut-on faire mieux que la force
brute (lien avec le phénotypage
haut débit : complémentaire ou
aveu d'échec de démarches
formalisées?)

**Merci d'avance de vos
commentaires...**