

Les outils de la génomique pour la sélection assistée par marqueurs chez le ray-grass anglais

Philippe Barre

INRA Centre Poitou-Charentes

UR4 : Unité de Recherche Pluridisciplinaire, Prairies et Plantes Fourragères
(URP^{3F})



Sommaire

É Introduction

É Etude de la diversité disponible pour la
sélection

É Identification de QTL pour la construction de
génotypes

É Sélection génomique

É Conclusion

Contexte

- “ Les prairies représentent **45% de la SAU** française (14 Millions dha) rôle majeur dans la **alimentation** des herbivores et dans d'autres services éco-systémiques (qualité de l'eau, stockage de C_o)
- “ Une partie des prairies utilisée dans les systèmes prairiaux les plus intensifs est semée régulièrement: **prairies temporaires** recouvrant 2,7 Millions dha
- “ Les prairies temporaires sont semées avec du **matériel sélectionné** pour différents caractères dont **le rendement, la résistance aux maladies, la qualité, la pérennité**À



Contexte

“ Le **ray-grass anglais** (*Lolium perenne* L.) est la espèce de graminées fourragères la plus semée dans les régions tempérées



“ C'est une espèce :

- diploïde **2n=14** avec un génome de taille moyenne 1C=2.3 Gb (séquençage en cours)
- très polymorphe : 1 SNP (Single Nucleotide Polymorphism) / 15-30 bases
- présentant un **système de auto-incompatibilité** et une forte **dépression de consanguinité**

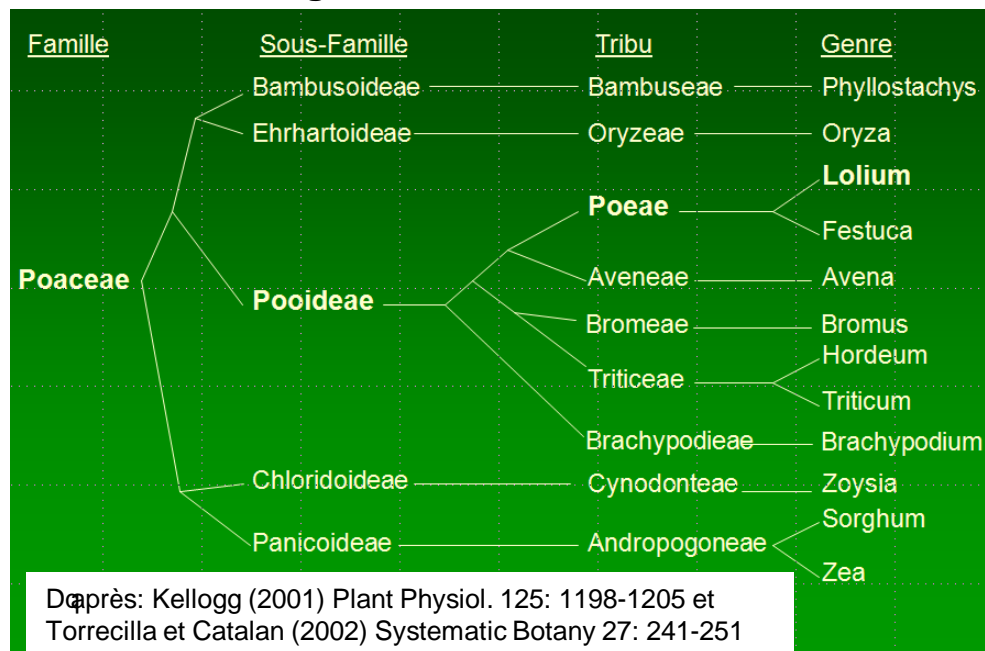
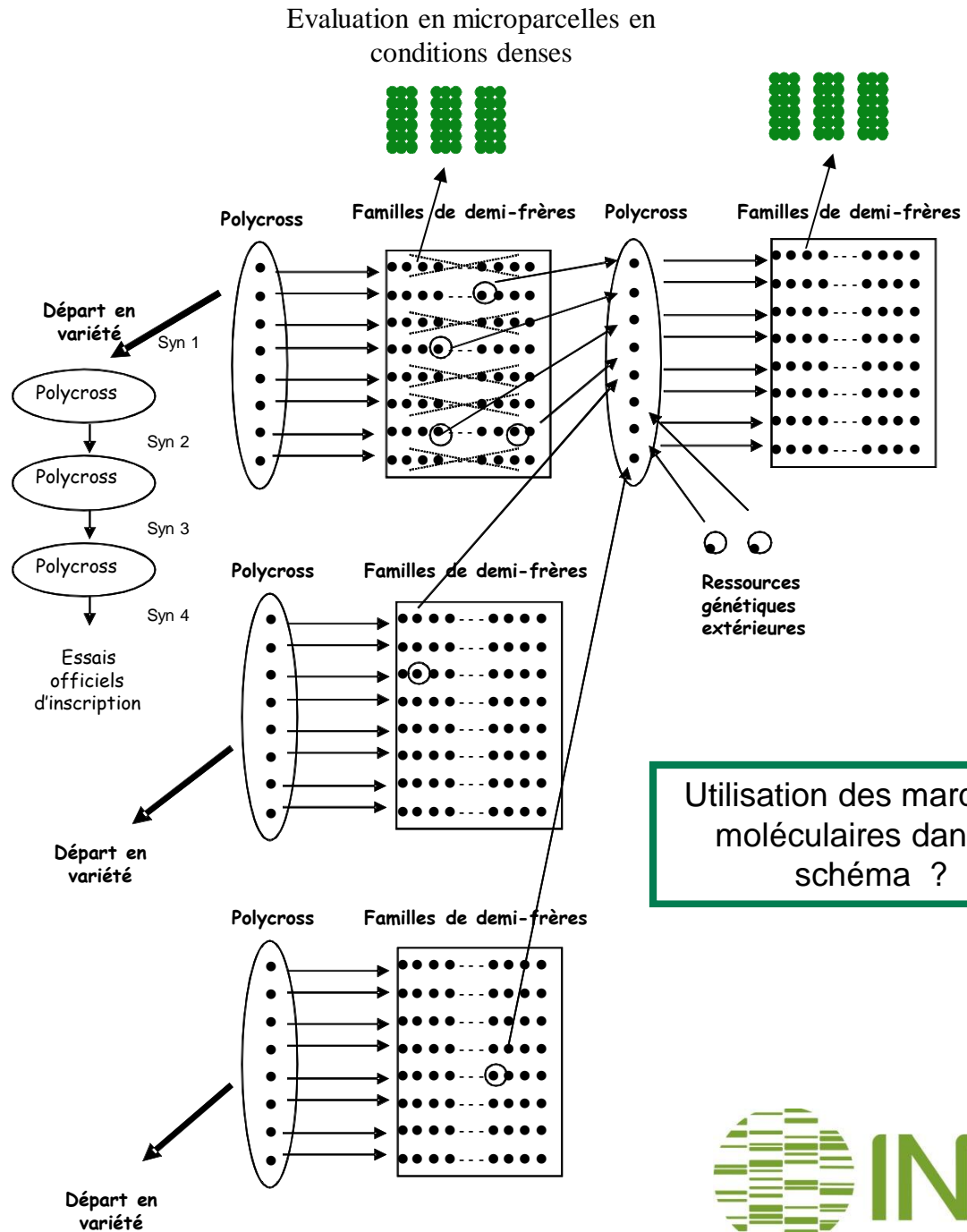


Schéma de création variétale du ray-grass anglais



Variétés
synthétiques



Marqueurs disponibles

É Marqueurs dominants révélés en masse:

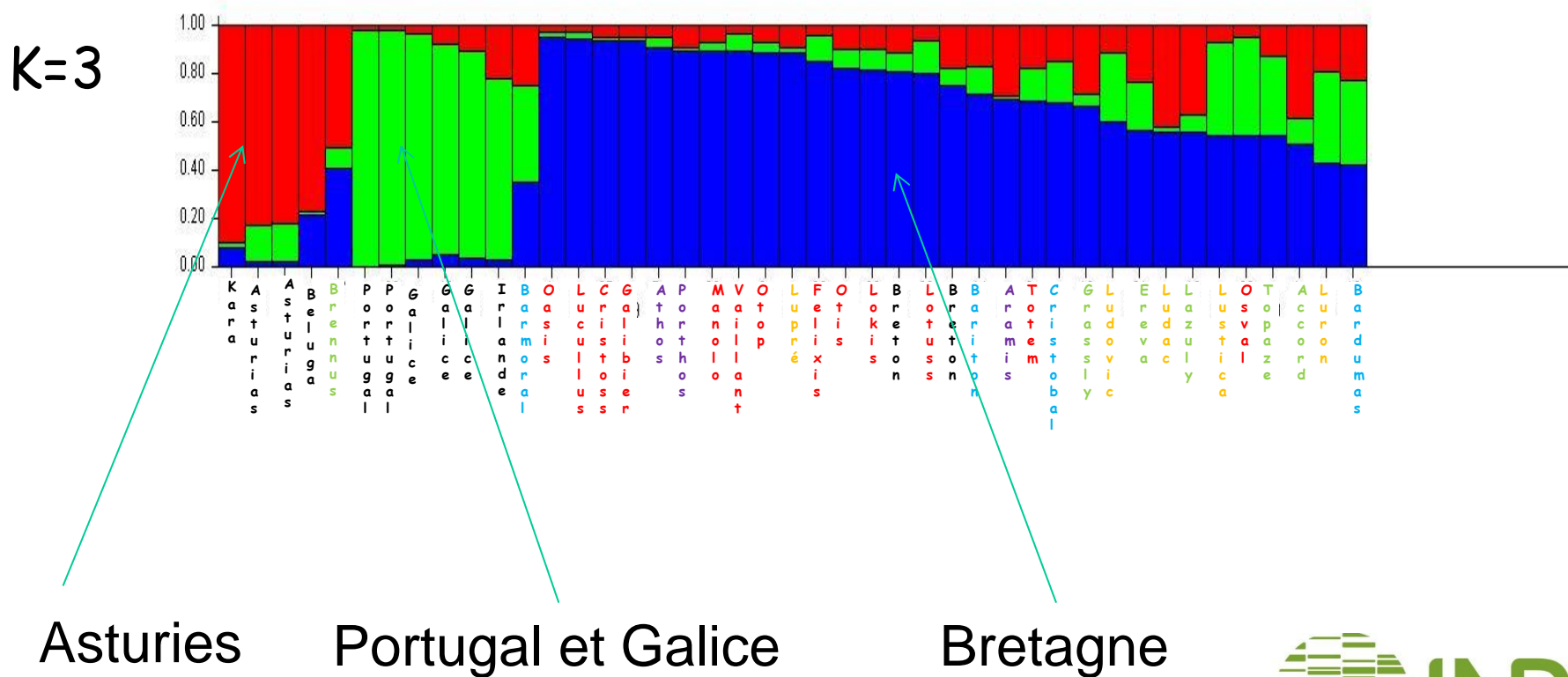
- ó AFLP: amplified fragment length polymorphism
- ó Dart : diversity array technology (plateforme CIRAD Montpellier)

É Marqueurs co-dominants:

- ó Microsatellites
- ó Marqueurs SNP révélés par différentes techniques
- ó Génotypage par séquençage (GBS)

Etude de la diversité disponible pour la sélection

É Exemple : dactyle marqueurs AFLP



Identification de QTL pour la construction de génotypes

Objectif : augmenter la fréquence des allèles favorables aux QTL

É Populations bi-parentales

É Populations multi-parentales connectées

É Populations dont le pédigrée est inconnu

ó Étude d'association « gènes candidats »

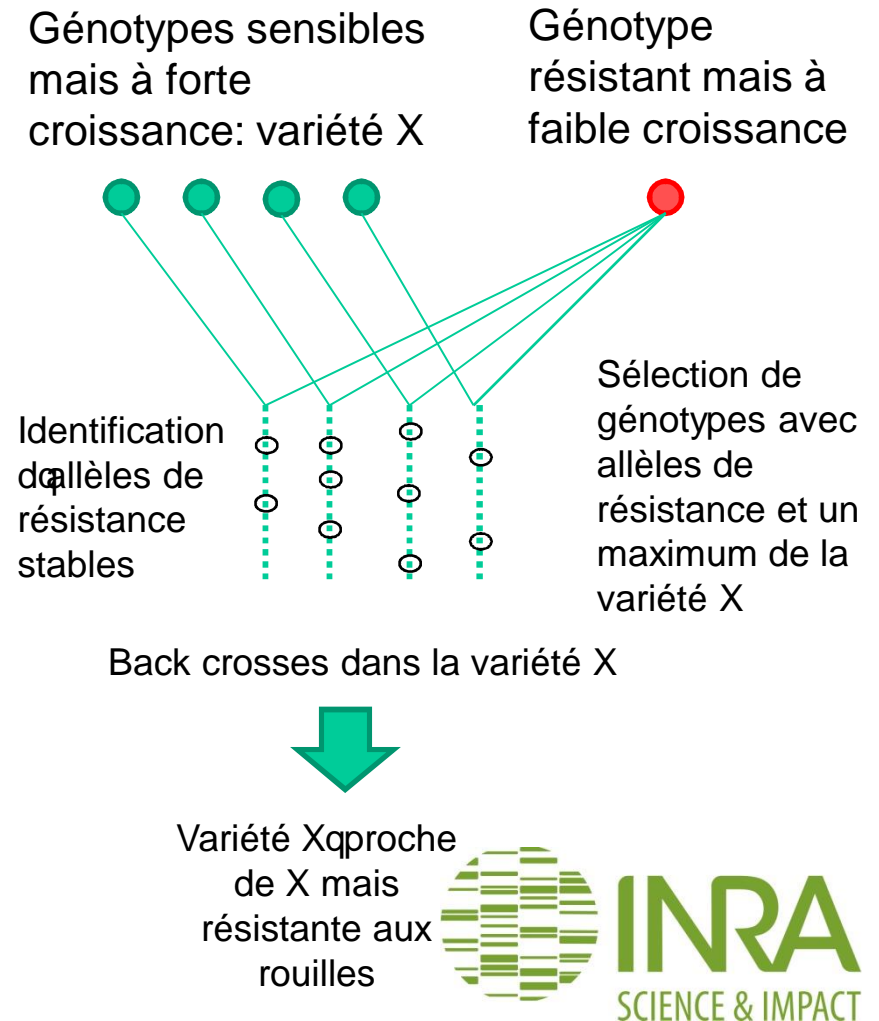
ó Etude d'association sur l'ensemble de génome:

GWAS Genome Wide Association Study

Identification de régions du génome impliquées dans les variations phénotypiques (QTL)

Utilisation de populations biparentales créées spécifiquement

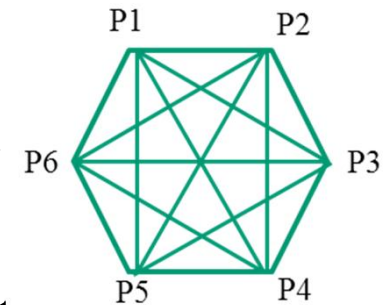
- Avantage : déséquilibre de liaison important
- Exemple: introgression d'allèles de résistance aux rouilles dans des variétés de ray-grass d'Italie
- Limites: uniquement deux parents, réalisation des croisements, difficiles pour des caractères très multigéniques comme le rendement



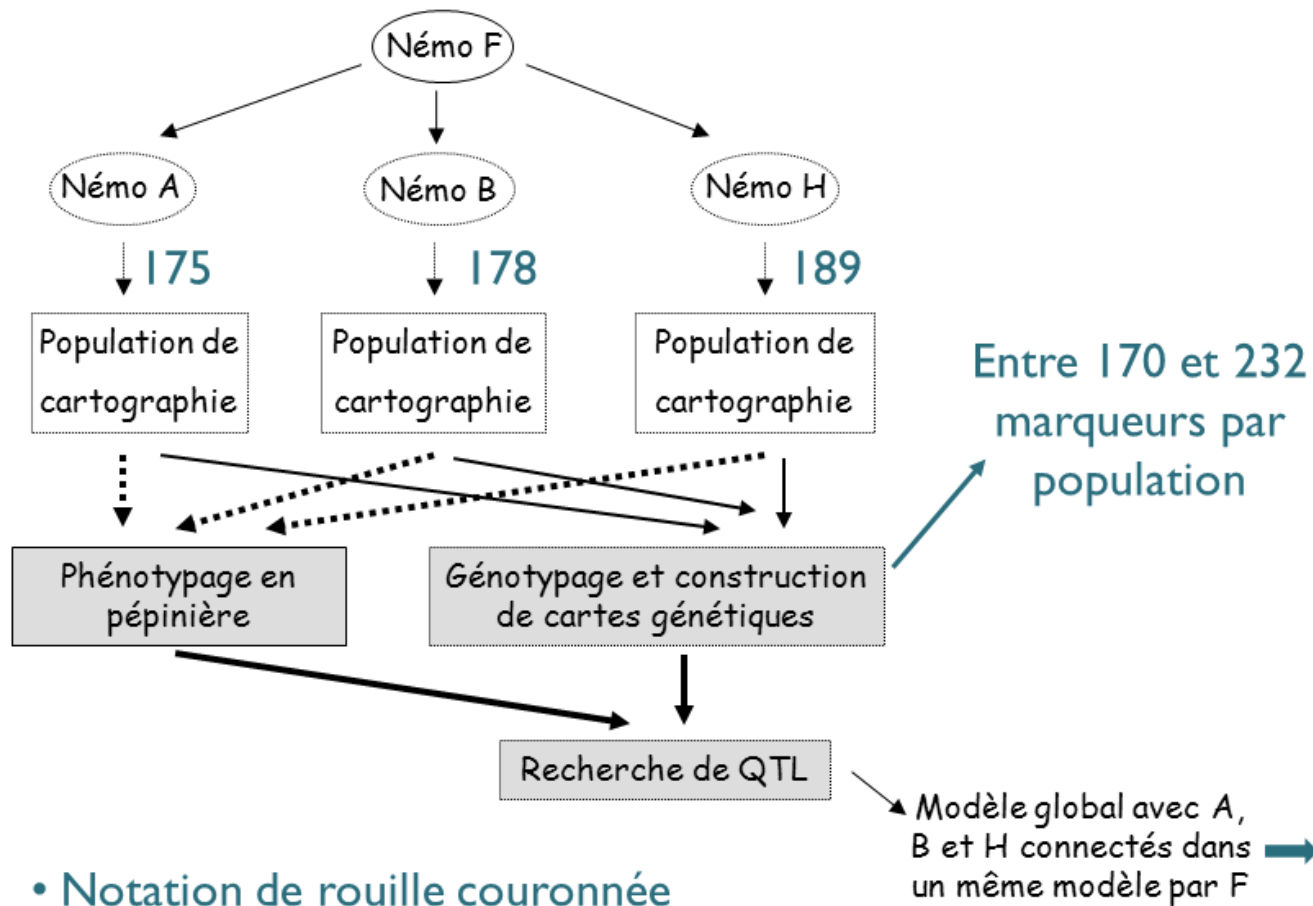
Identification de régions du génome impliquées dans les variations phénotypiques (QTL)

“ Utilisation des descendance de polycross classiquement réalisées par les sélectionneurs

- . Les descendance de polycross sont des **familles de pleins-frères connectées** qui est possible d'identifier grâce aux marqueurs moléculaires : tests de paternité
- . Ces familles ont un **déséquilibre de liaison important**
- . **Variabilité** explorée plus **importante** que dans les populations bi-parentales: dépend du nombre de parents du polycross et de leur apparentement
- . Exemple : hauteur étirée chez le ray-grass anglais (Thèse Laurence Pauly bourse CIFRE Jouffray-Drillaud)



Exemple 1 de populations connectées



- Notation de rouille couronnée
- Paramètres de croissance de plante

Analyse multipopulation . MCQTL

Brigitte Mangin

Némo A x Némo F
ab x cd

a vs b c vs d
i = a ou b j = c ou d

Némo B x Némo F
ef x cd

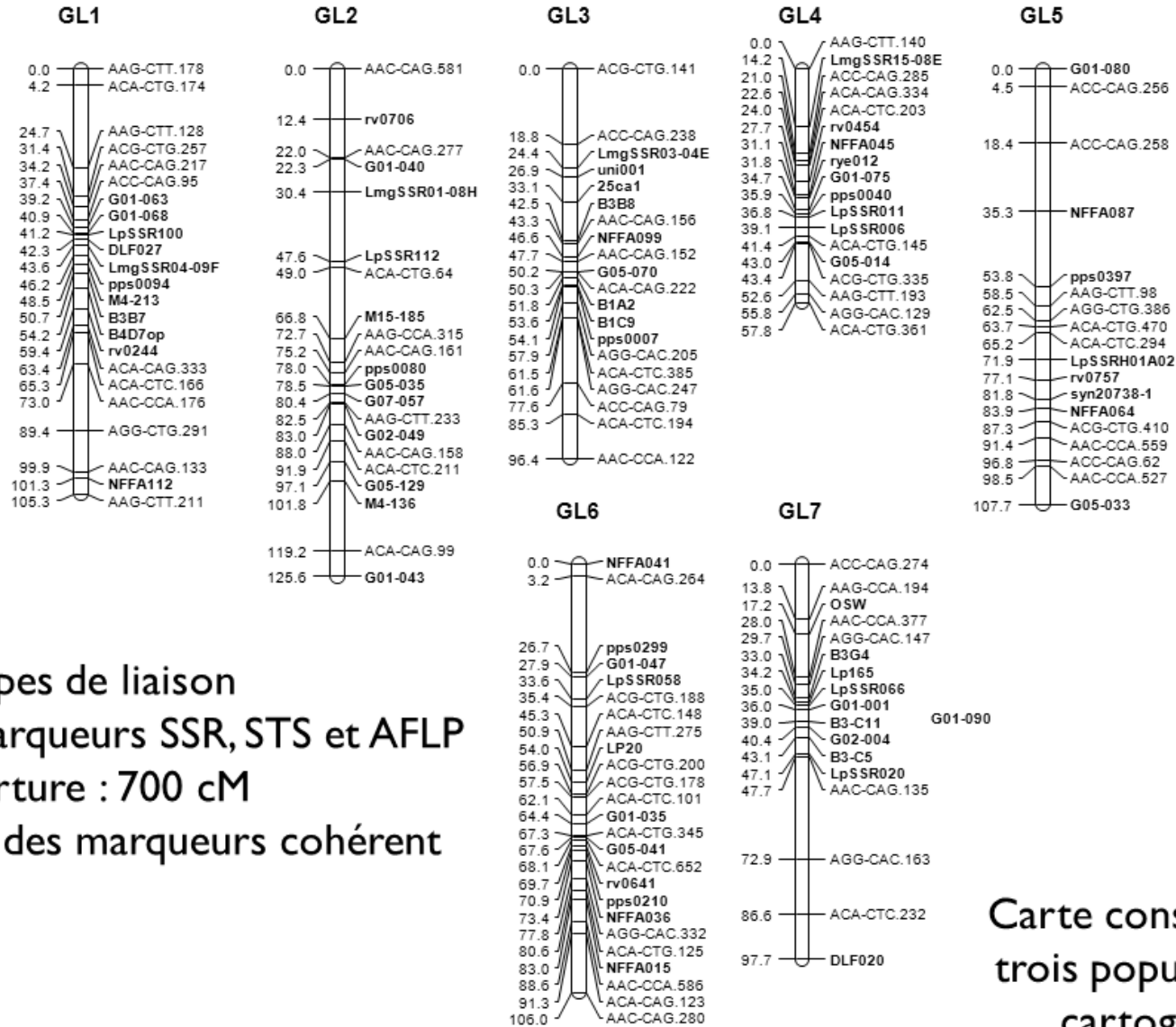
e vs f c vs d
m = e ou f n = c ou d

Némo H x Némo F
gh x cd

g vs h c vs d
o = g ou h p = c ou d

◦ Connectées : $\alpha_j = \alpha_n = \alpha_p = \alpha_t$ t = c ou d

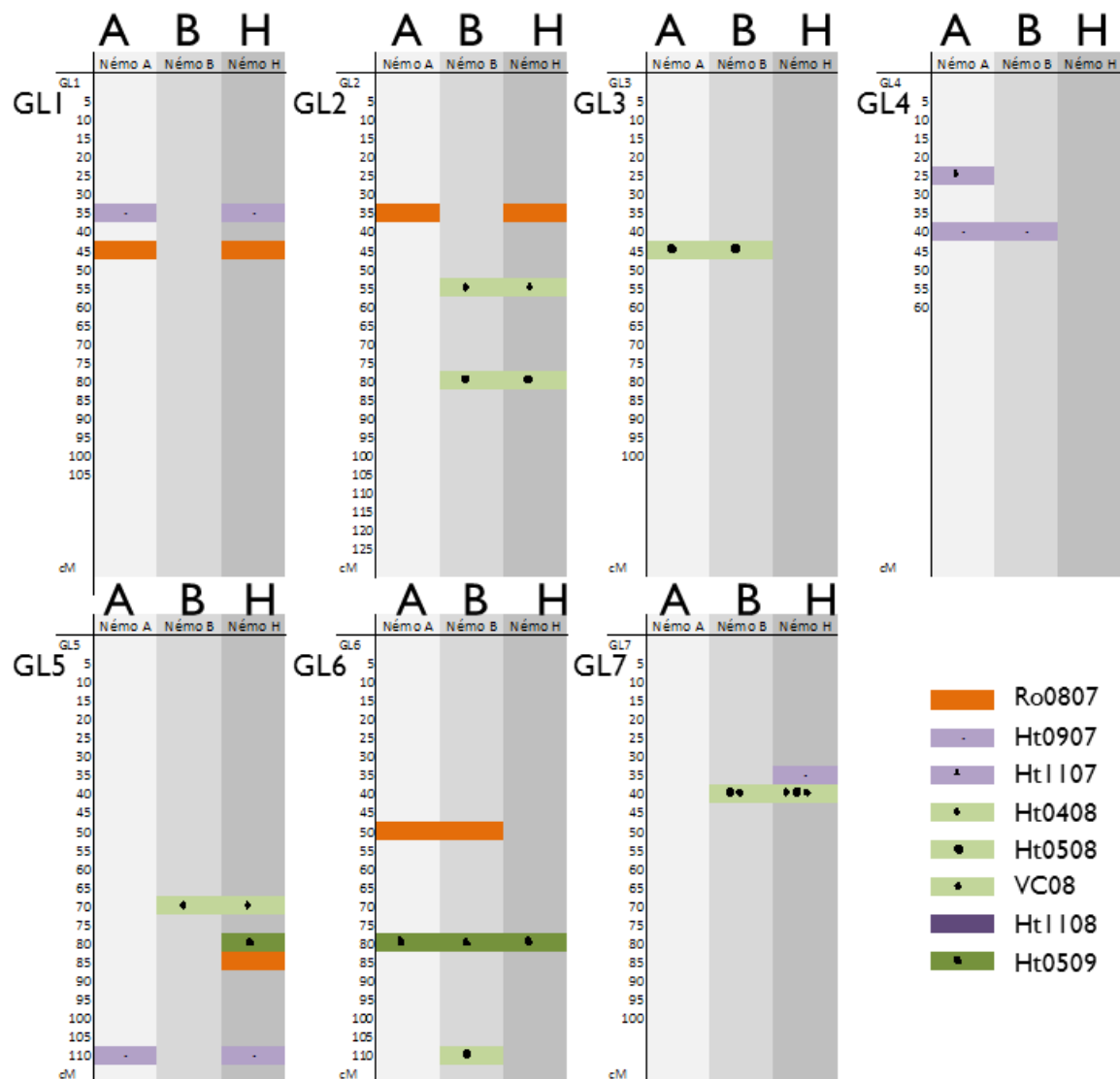
$$P_{kc} = \mu_c + \underbrace{\left. \begin{array}{l} \alpha_i \} A \\ \alpha_m \} B \\ \alpha_o \} H \end{array} \right\} F}_{\text{Effets additifs}} + \text{effets de dominance} + \varepsilon$$



- 7 groupes de liaison
- 142 marqueurs SSR, STS et AFLP
- Couverture : 700 cM
- Ordre des marqueurs cohérent

Carte consensus des
trois populations de
cartographie

QTL chez Némo A, Némo B et Némo H en analyse multipopulations connectées



- Des allèles favorables dans différents parents
- Des effets peu répétables dans le temps

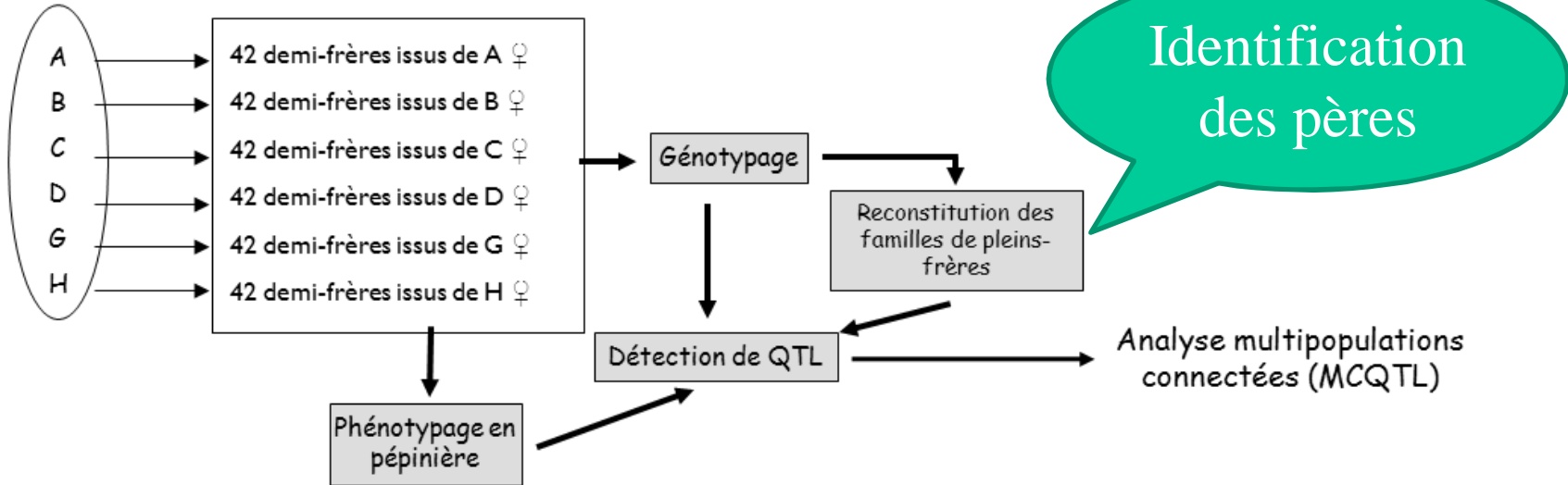
QTL chez Némo F en analyse simple population : vérification de l'hypothèse effet de F identique dans tous les croisements



- Un QTL de résistance aux rouilles à effet fort et stable entre croisements
- Des QTL non stables à effet faible en général
- Cependant l'analyse en populations connectées permet d'identifier le plus de QTL

Exemple 2 de populations connectées

Polycross



Nombre d'individus par famille de pleins-frères					Nombre d'individus par famille de demi-frères
Némo A	Némo B	Némo C	Némo D	Némo G	

Némo A						84
Némo B	20					55
Némo C	5	3				48
Némo D	27	5	10			67
Némo G	23	22	2	16		70
Némo H	9	5	28	9	7	58

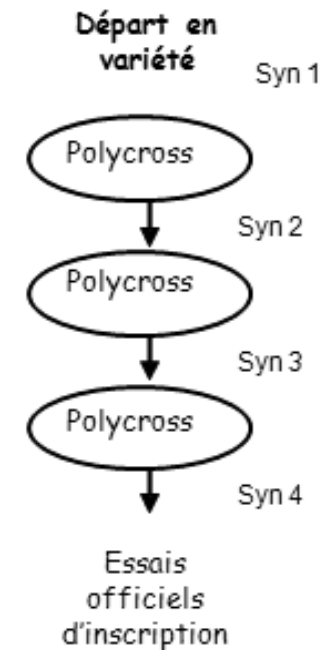
Détection de QTL dans la descendance du polycross grâce à l'analyse multipopulations connectées de MCQTL

Caractère	GL	Marqueur	Position sur la carte consensus (cM)	R ² (%)	Effets additifs					
					Némo A	Némo B	Némo C	Némo D	Némo G	Némo H
Ht0309	3	pps0007	29,3	12,0	-1,8	4,5	5,2	4,6	-9,5	1,9
Ht0409_1	3	pps0007	29,3	14,0	-1,7	0,9	5,8	4,8	-5,5	1,0
Ht0409_2	3	pps0007	29,3	15,6	-1,2	3,0	8,8	8,3	-9,7	3,1
Ht0309	4	LpSSR011	30,8	20,1	-3,8	7,4	10,7	-1,2	3,4	10,8
Ht0409_1	4	G01-073	28,8	19,2	-3,3	4,1	5,7	0,1	4,0	6,0
Ht0409_2	4	G01-075	27,8	17,3	-4,6	4,5	9,6	1,5	5,6	10,6
Ht0409_3	4	pps0040	28,2	17,3	-6,6	4,3	7,4	1,2	6,7	8,7
Ht0409_4	4	LpSSR011	31,7	16,1	-5,3	11,8	7,8	0,8	4,0	9,1
Ht0409_5	4	LpSSR011	30,8	22,1	-6,6	12,4	11,7	-2,6	8,4	11,6
VC09	4	G05-014	33,8	12,3	-0,05	0,08	0,11	-0,04	0,02	0,06

Identification de régions du génome impliquées dans les variations phénotypiques (QTL)

Objectif : augmenter la fréquence des allèles favorables aux QTL

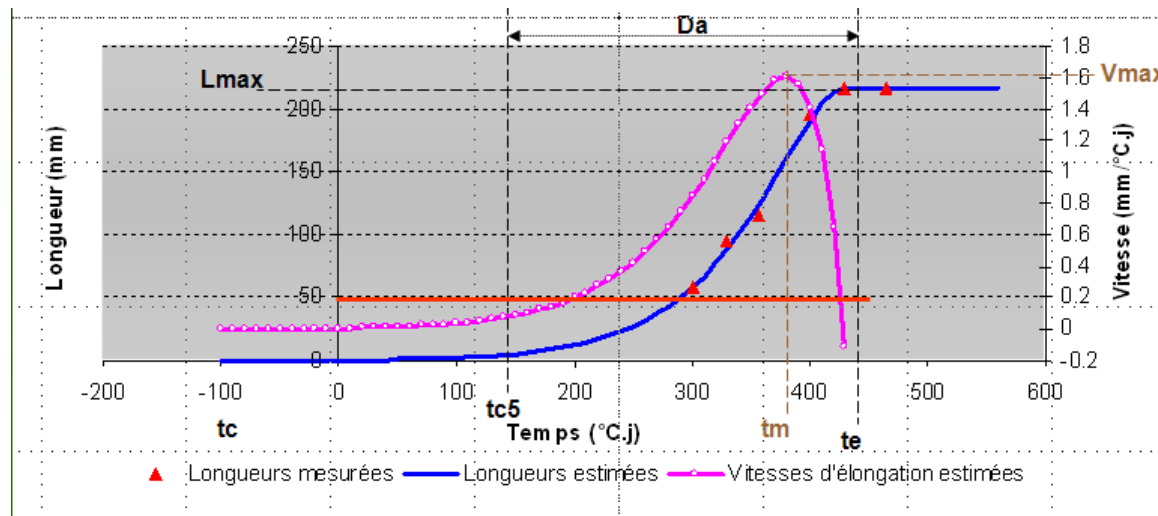
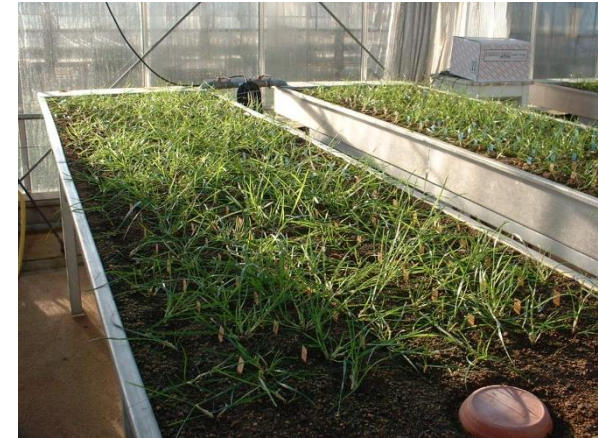
- “ Utilisation de populations bi-parentales créées spécifiquement
- “ Utilisation des descendance de polycross classiquement réalisées par les sélectionneurs: analyse en multipopulations connectées
- “ Utilisation de variétés synthétiques: étude d'association
 - . Variabilité importante et populations non structurées
 - . Déséquilibre de liaison faible:
 - “ Analyse du génome avec énormément de marqueurs
 - “ Analyses ciblées sur des régions d'intérêt: « gènes candidats »



Exemple: étude de l'association entre la croissance foliaire et le gène GAI (Gibberelic Acid Insensitive)

Thèse de Jérôme Auzanneau

- 186 génotypes de la variété Herbie
- cinétique de croissance foliaire au printemps et à l'automne
- séquençage d'une partie du gène



Absence de structuration

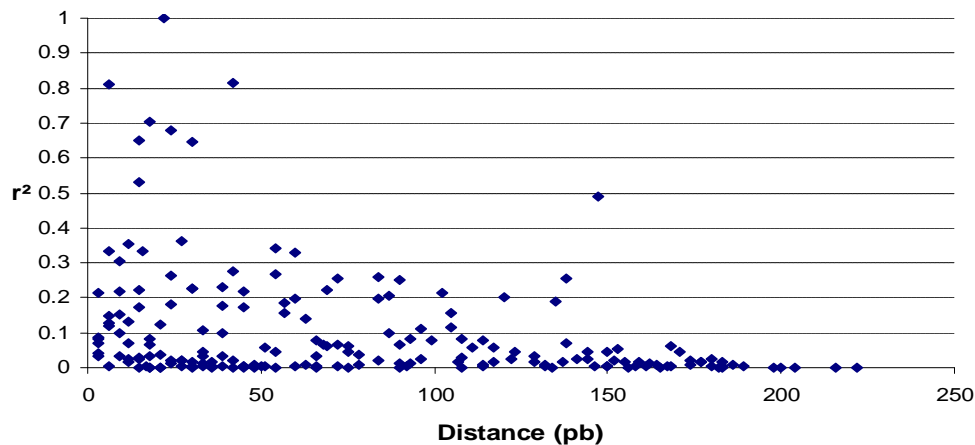
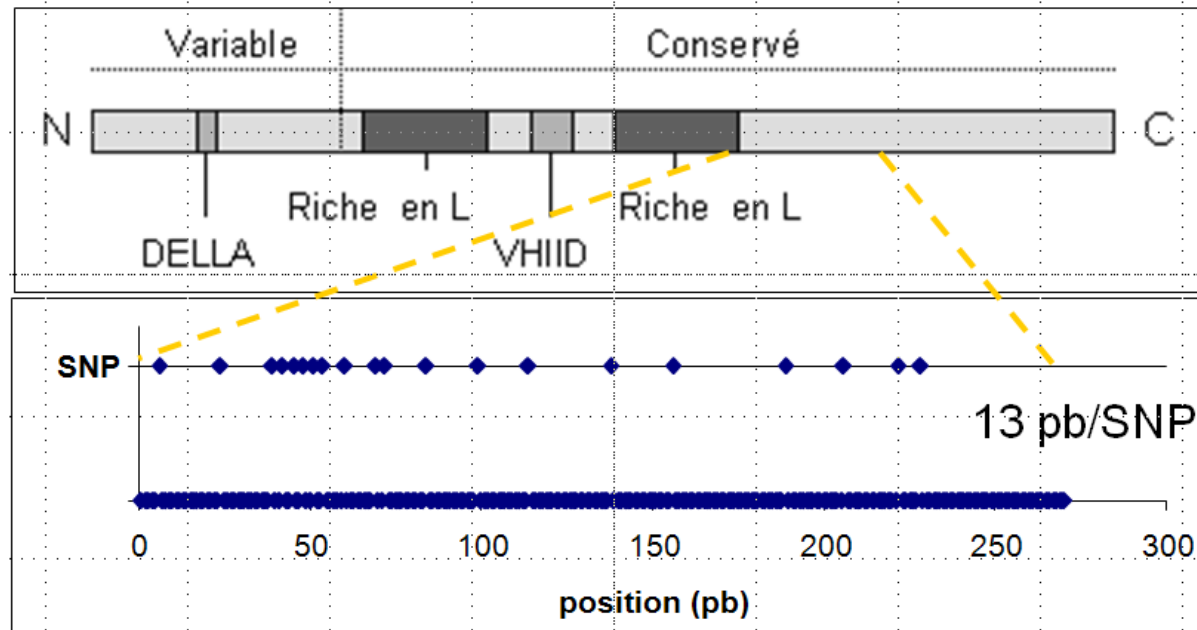
- É Utilisation de marqueurs répartis sur le génome pour vérifier l'absence de structuration de la population étudiée
- É Essentiel pour éviter de détecter des faux positifs
- É Si il existe une structuration, il faut la prendre en compte dans le modèle
- É Possibilité de prendre aussi en compte l'apparementement entre individus

Effet génétique et variabilité de la croissance foliaire

Variable	Printemps		Automne	
	Effet géno. (p)	Héritabilité individuelle	Effet géno. (p)	Héritabilité individuelle
Lmax	0,0001	0,77	0,0001	0,64
Vmax	0,0001	0,74	0,0001	0,67
Da	0,0001	0,46	0,0001	0,39

Variable	Max	Min	Moy
Her4 Lmax (mm)	450	126	301
Her4 Vmax (mm/°Cj)	2,56	0,84	1.64
Her4 Da (°Cj)	401	187	267

Variabilité du gène GAI et DL chez Herbie



- fort polymorphisme : 20 SNP
- décroissance du déséquilibre de liaison très rapide

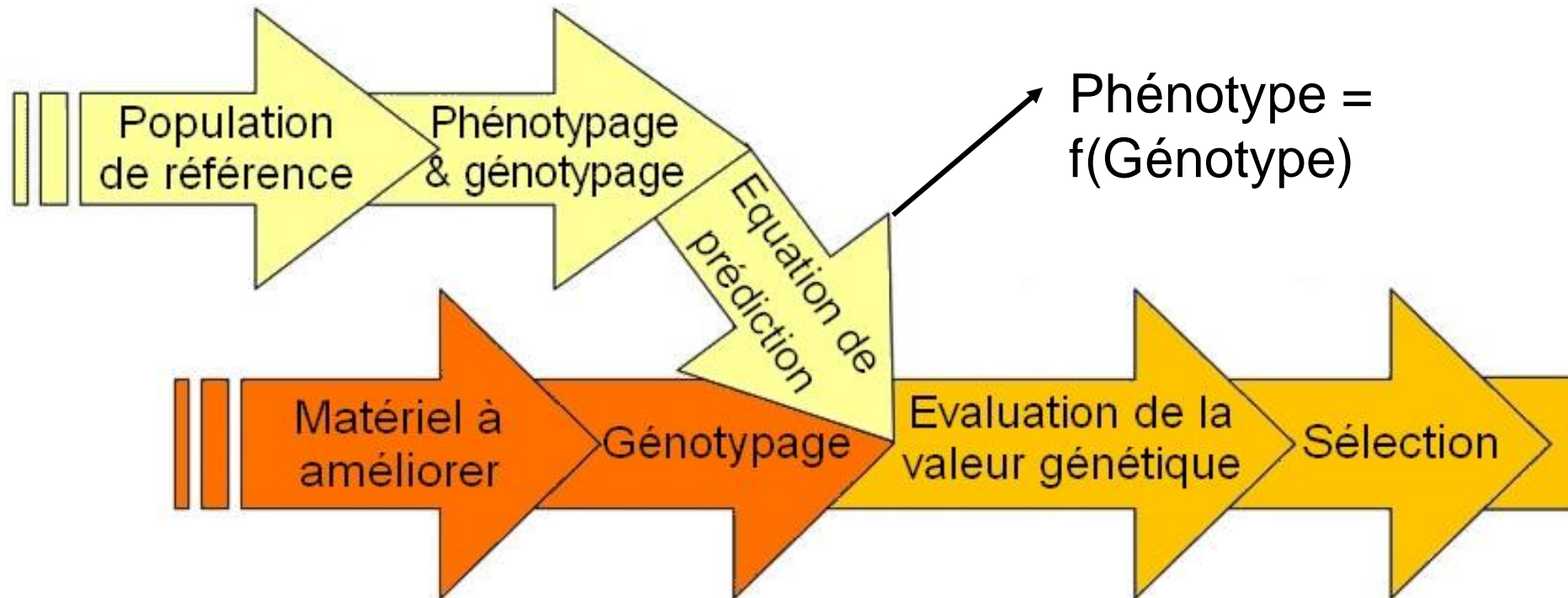
Régression linéaire multiple sur les 20 SNP

Printemps				Automne			
Variables	SNP	R ² partiel	R ² global	Variables	SNP	R ² partiel	R ² global
Lmax (mm)	GAI069	0,0502	0,0934	Lmax	GAI069	0,038	0,1106
	GAI060	0,0255		(mm)	GAI099	0,0295	
	GAI039	0,0177		GAI048	0,0252		
			GAI222	0,018			
Vmax (mm/°Cj)	GAI069	0,0407	0,1008	Vmax	GAI069	0,0418	0,1358
	GAI099	0,0233		(mm/°Cj)	GAI048	0,0219	
	GAI206	0,0185		GAI114	0,018		
GAI189	0,0183	GAI156	0,0174				
			GAI099	0,0122			
			GAI189	0,0245			
Da (°Cj)	GAI069	0,0187	0,0755	Da	GAI051	0,0341	0,0838
	GAI060	0,0164		(°Cj)	GAI069	0,0225	
	GAI039	0,0159		GAI114	0,0144		
	GAI048	0,0124		GAI048	0,0128		

Un SNP explique jusqu'à 5% de la variation de longueur de limbe mais

Sélection génomique

Principe de la sélection génomique



D'après Heffner et al. (2009)

Hypothèse : Équation de prédiction valable dans la population à améliorer

Objectif

Évaluation de la précision de prédiction pour deux caractères à héritabilité contrastées dans une population synthétique de ray-grass anglais à base génétique étroite après trois générations d'inter-croisements
Stage M2 Camille Gréard

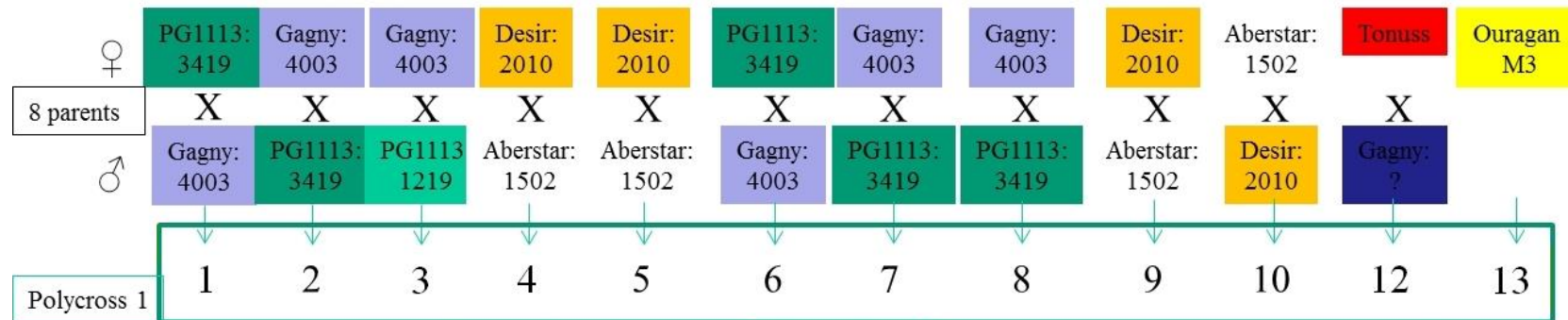
Caractères étudiés :

- **Longueur de feuilles** : Lié au rendement
héritabilité sens large individuelle : 0,45-0,7
- **Date d'épiaison** : Catégorisation de la précocité
héritabilité sens large individuelle : 0,7- 0,95

Population à base étroite (peu de fondateurs très différents) : créer du **déséquilibre de liaison** physique

Population créée par GIE-Grass

Matériel et méthodes : matériel végétal



M1

Polycross 2

M2

Polycross 3

M3

GBS (17 878 marqueurs)

← Génotypage

450 plantes

→ Phénotypage

Longueur de feuilles

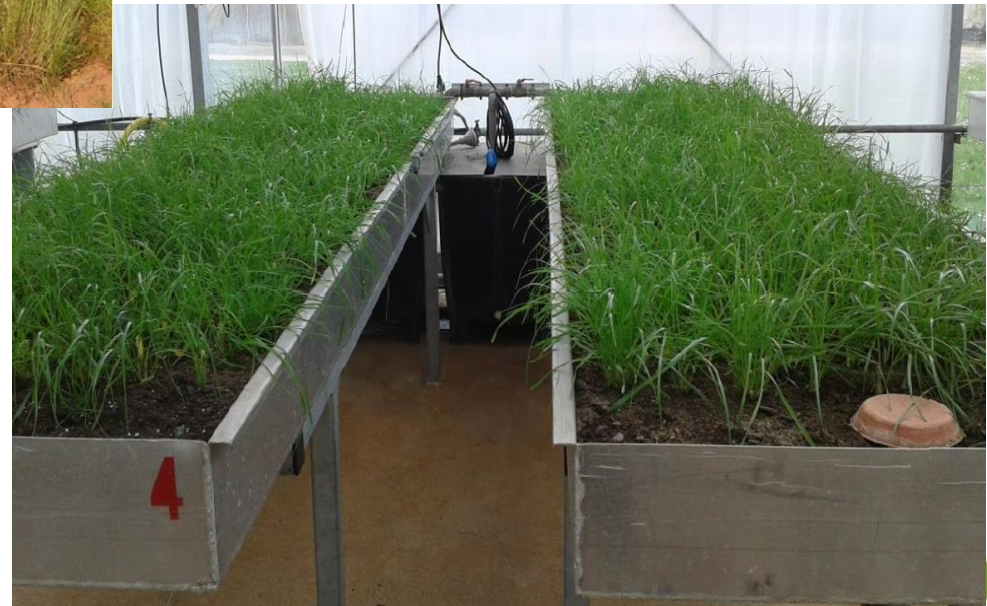
Date d'épiaison

Matériel et méthodes : phénotypage

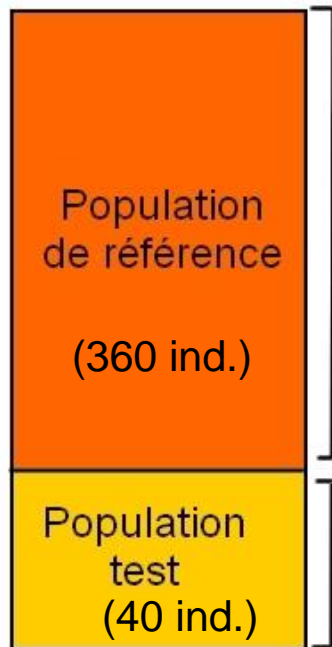


Date d'épiaison : Pépinière
Installation : 16/04/2014
Plante épiée : 3 épis visibles
Mesures du 20 avril au 31 mai 2015

Longueur de feuilles : serre
Installation : 17/12/2014
Feuille mature : ligule visible
Mesures du 23 février au 30 mars 2015



Matériel et méthodes : Modèles



Calibration de
l'équation de prédiction



Prédiction de la
valeur génétique

Modèle:

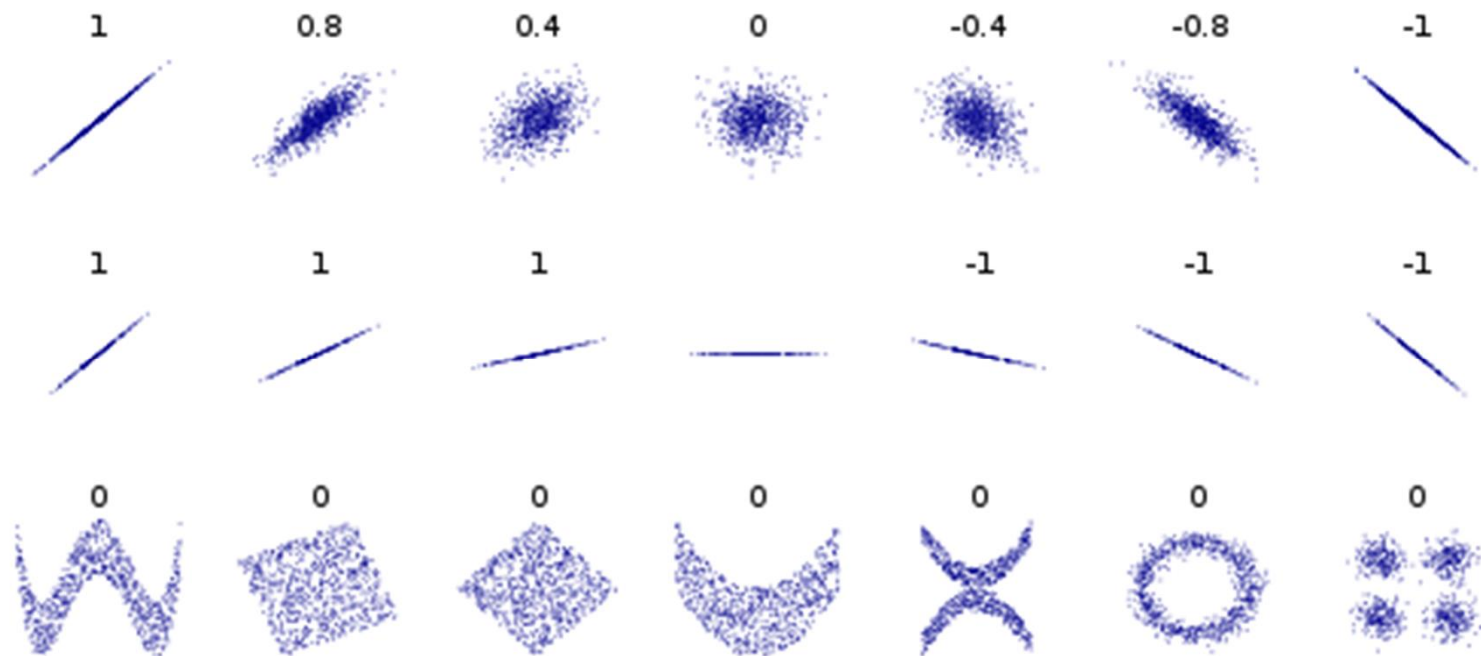
$$\text{Phéno} = \text{Moy} + \text{somme}(\text{Génotype} * \text{effet}) +$$

On estime les effets des allèles pour tous les locus sur une population de référence

On utilise les valeurs estimées sur la population de référence pour calculer les valeurs prédites de la population test

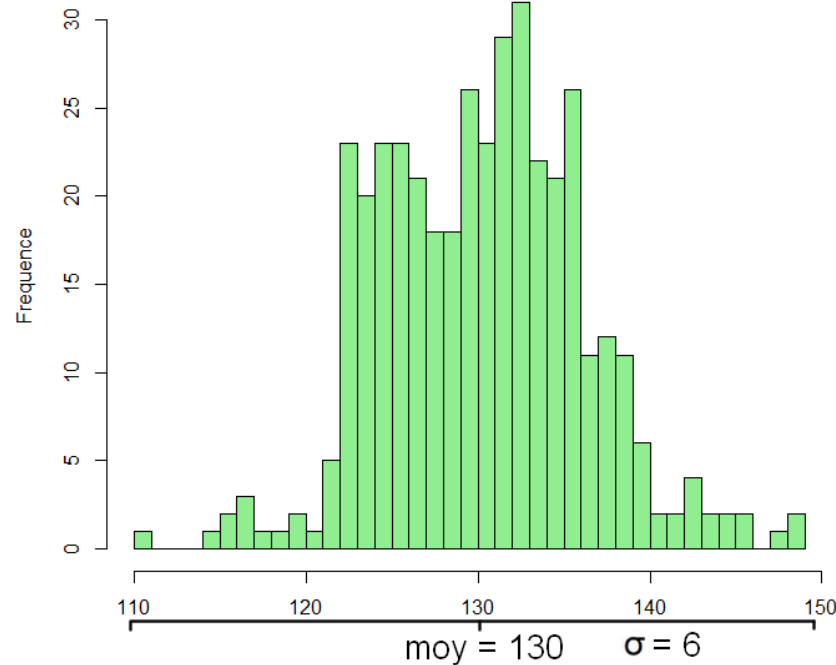
Matériel et méthodes : la précision

Précision de prédiction :
Coefficient de corrélation de Pearson entre
les valeurs prédites (G_p) et les valeurs phénotypiques (P)

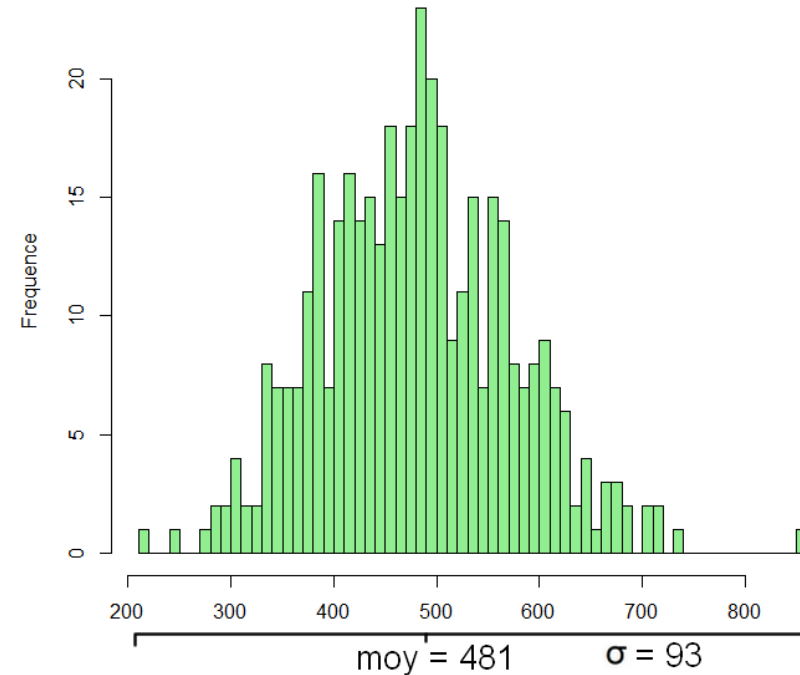


Résultats et discussion

Distribution des moyennes ajustées de la date d'épiaison (jours)



Distribution des moyennes ajustées de la longueur de feuille 4 (mm)



	Date d'épiaison	Longueur de feuille 4
Héritabilité individuelle	0,96	0,59
Héritabilité des moyennes génotypiques	0,99	0,81

Forte variabilité des caractères mesurés

Résultats et discussion

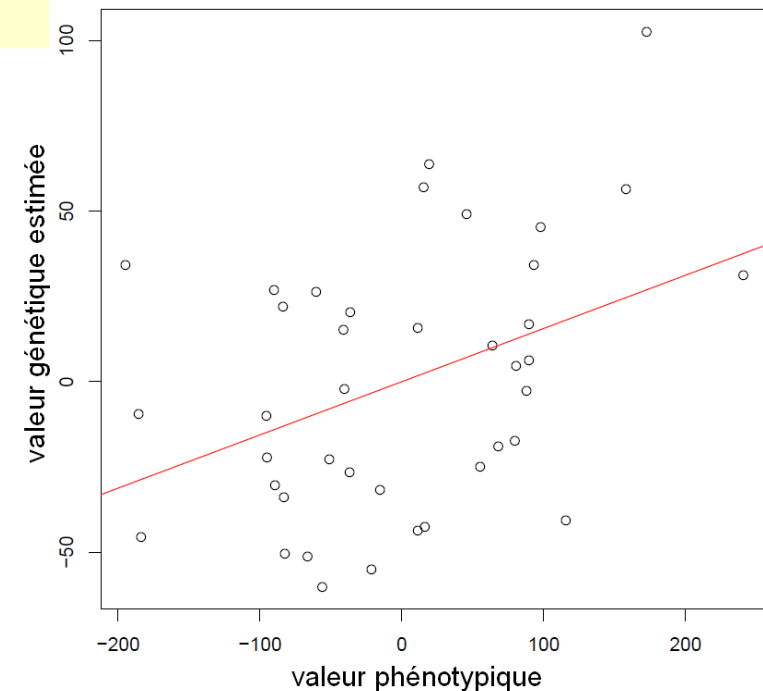
Précision de prédiction RR-Blup	Moyenne
Date d'épiaison	0,40
Longueur de feuille	0,22

La précision de prédiction suit l'héritabilité

Valeurs de précision pour d'autres espèces :

- Date d'épiaison (blé) : 0,40
- Hauteur de plante (riz) : 0,34

Exemple de tirage avec un coefficient de 0.4

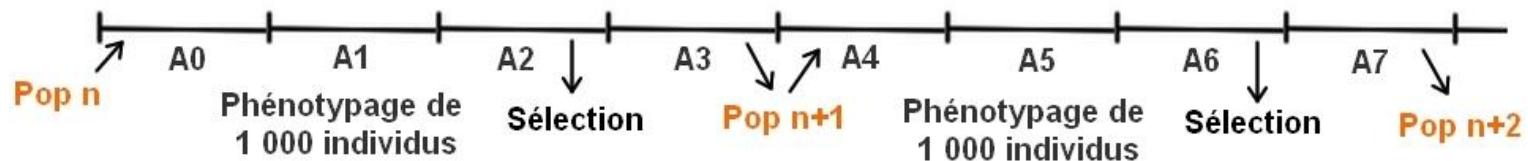


Résultats et discussion

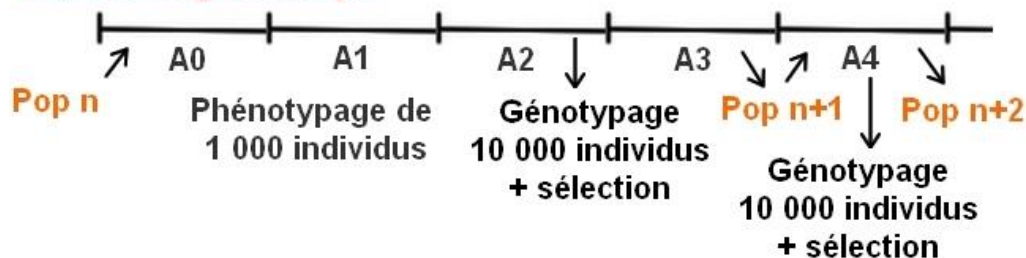
Progrès génétique en sélection sur le phénotype : $\Delta G = \frac{i \times h_{SL\ moy\ geno}^2 \times \sigma_p}{T}$

Progrès génétique en sélection génomique : $\Delta G = \frac{i \times cor(G_p, P) \times \sigma_p}{T}$

Sélection sur le phénotype



Sélection génomique






Pour avoir un **gain génétique** supérieur en sélection génomique :

- Précision de prédiction date d'épiaison > 0,44
- Précision de prédiction longueur de feuille > 0,36

Conclusion et perspectives

Sélection génomique

-  Les valeurs d'héritabilité correspondent aux valeurs attendues
-  Caractère plus héritable mieux prédit
-  Amélioration de la précision de prédiction
 - Diminuer le nombre de données manquantes en génotypage
 - Augmenter la densité de marquage
 - Prendre en compte les haplotypes aux loci plutôt que chaque marqueur individuellement

Conclusion

- ” L'utilisation des marqueurs chez le ray-grass anglais:
- . **Gestion de la diversité**
 - . **Identification des pères** : sélection sur les deux parents permet de multiplier le gain génétique par deux
 - . **Introgression d'allèles favorables** dans du matériel élite pour des allèles avec des effets forts (QTL et étude de **association**)
 - . Début de la **sélection génomique** pour les caractères contrôlés par de nombreux gènes à effet faible

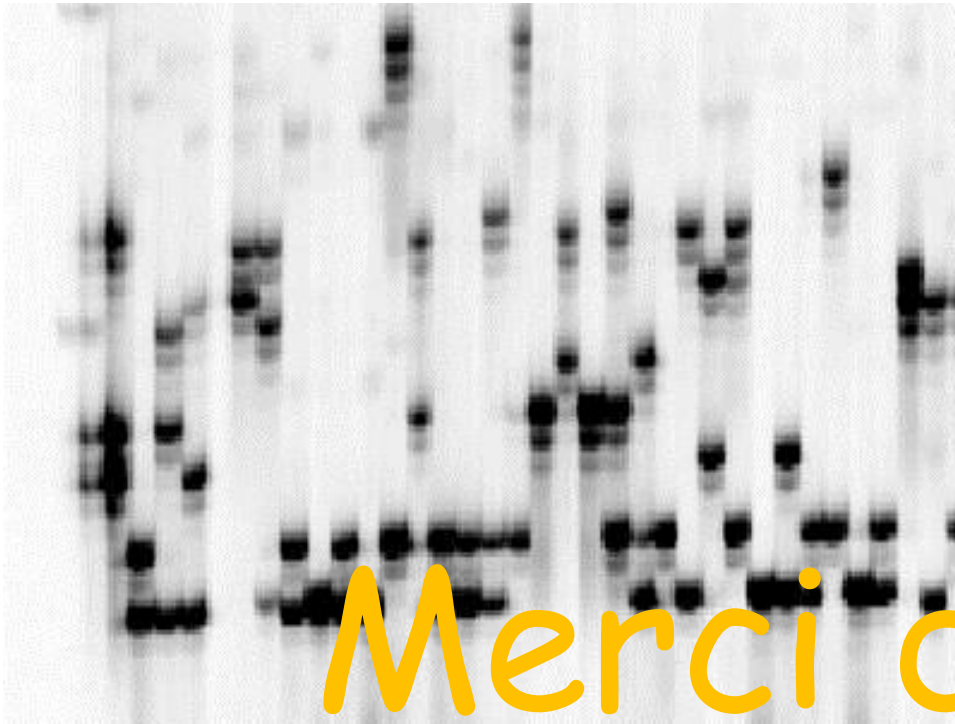
Conclusion

É L'utilisation des marqueurs va dépendre:

ó du coût par rapport aux bénéfices

ó de la précision de prédiction avec les marqueurs

ó de la facilité de mise en œuvre (personnes compétentes, marqueurs moléculaires disponibles, rapidité de génotypage et d'analyse)



Merci de votre
attention!



A screenshot of a bioinformatics software interface, likely a sequence alignment viewer. The window title is "[AU249224ref...]. X". The main area displays a sequence alignment for "AU249224ref" (Torben) against a reference sequence. The alignment shows a consensus sequence: GGAGAG - AACAAAGAT - AAATT - AGCAGTGAATCC - AAGC - A - TTACAGGTGCGGATCCAATTATTTTC - AA. Below this, multiple individual reads are shown, with some mismatches highlighted in red. A pink bar highlights the consensus sequence. On the right side, there is a "Read Mapping Settings" panel with various options: "Read layout" (Compactness: Packed), "Gather sequences at top" (checked), "Show sequence ends" (checked), "Show mismatches" (checked), "Packed read height: Medium", "Find Conflict" button, "Low coverage threshold: 8", "Find Low Coverage" button, "Sequence layout" (No spacing), "Numbers on sequences" (checked), "Relative to: 1", "Numbers on plus strand" (checked), "Lock top sequence" (checked), "Hide labels" (unchecked), and "Lock labels" (checked).