



Génétique
Quantitative
et Évolution
Le Moulon

INRAE

La sélection génomique chez le Maïs

L. Moreau, A. Allier, H. Giraud, A. Lorenzi, S. Rio, A. Seye, A. Charcosset



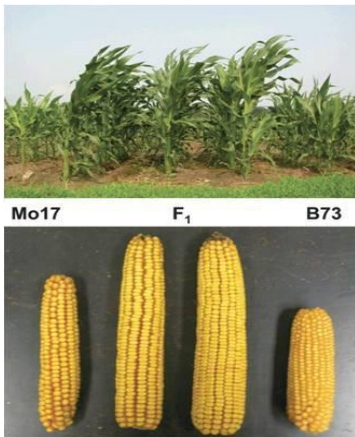
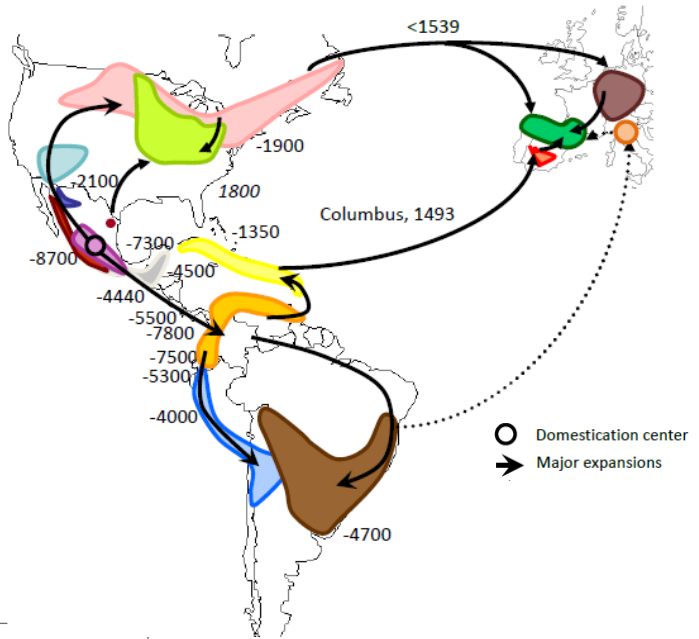


Plan

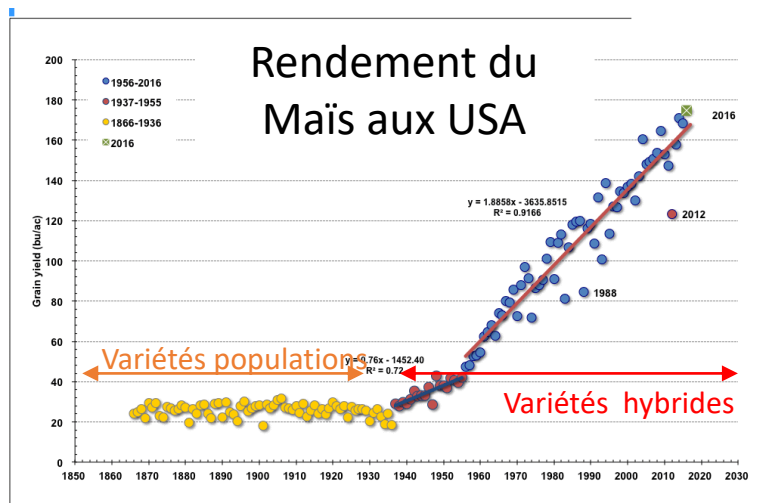
1. Éléments de contexte sur la sélection du maïs
2. De la Sélection Assistée par Marqueurs à la sélection génomique chez le maïs
3. La sélection génomique pour optimiser la sélection de lignées améliorées
4. La sélection génomique pour revisiter la sélection hybride
5. La sélection génomique pour mieux gérer la diversité génétique
6. Conclusions et Perspectives : la sélection génomique à l'aire de la sélection 4.0 ?

L'amélioration du maïs: éléments de contexte

- Domestication au Mexique puis adaptation aux zones tempérées
- Espèce allogame, fort hétérosis
- Années 50 : révolution liée au passage populations -> hybrides simples
- Formation de groupes « hétérotiques »
- Secteur privé très compétitif, lignées élites non partagées (sauf « ex pvp »)

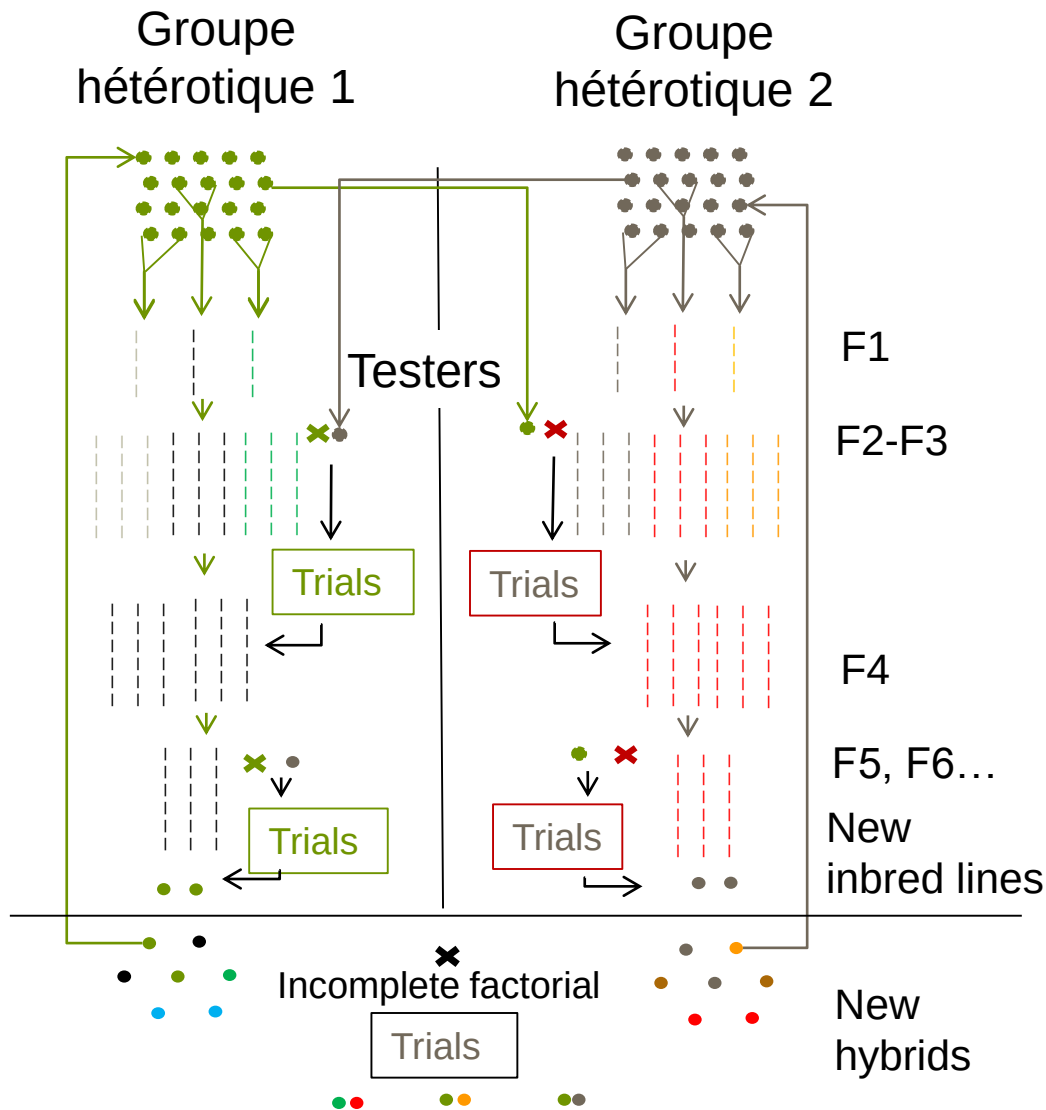


Springer and Stupar, 2007



⇒ Variétés populations :
réservoir de variabilité
génétique partiellement
exploitée lors du passage
aux variétés hybrides

Schéma de sélection hybride du maïs



Sélection **intra-** puis **inter-** groupes:

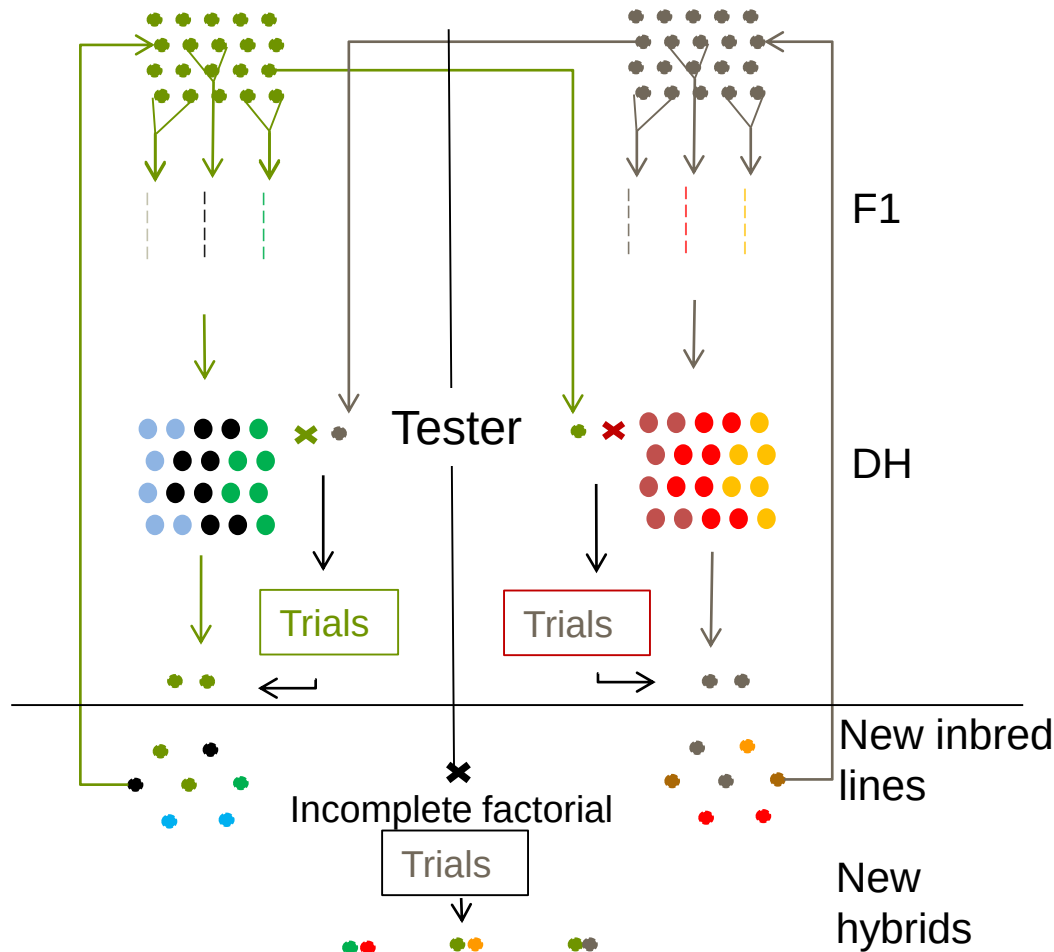
1. **Croisements intra-groupe** et création de familles en ségrégation

2. **Evaluation de la valeur en croisement** avec testeur(s) du groupe complémentaire
Aptitude Générale à la Combinaison (AGC)

3. **Evaluation des meilleures combinaisons hybrides inter-groupe**

$$AGC_i + AGC_j + ASC_{ij}$$

Schéma de sélection hybride du maïs



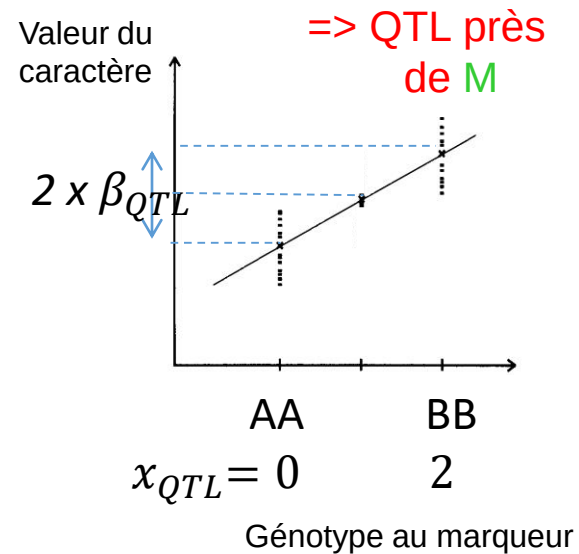
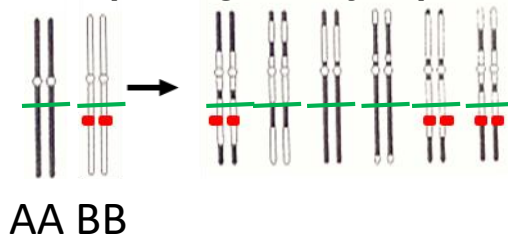
L' utilisation massive de **l'haplo-diploïdisation** permet de générer à partir des hybrids F1 des centaines de lignées par cycle.

⇒ **Processus long** (8 ans si SSD, 5 ans si HD / cycle), **couteux** (phénotypage), **valorisant peu la complémentarité** des parents de chaque groupe à un stade précoce

« Sélection Génomique » ou « Sélection Génome-wide »

Le marquage moléculaire (années 90) a ouvert la voie vers la localisation des locus impliqués dans la variation des caractères (QTL) et la calibration de modèles permettant de prédire la valeur génétique d'individus non évalués (Sélection Assistée par Marqueurs).

Détection QTL intra-famille (Linkage Analysis)



Malgré quelques succès (ex Eathington et al., 2007), intérêt limité de la SAM lié aux limites statistiques de la détection de QTL (déterminisme complexe, faible puissance, effets surestimés...) chez le maïs

Sélection assistée par marqueurs post QTL (SAM)

$$G = M = \sum_{QTL} \beta_{QTL} x_{QTL}$$

« Sélection Génomique » ou « Sélection Génome-wide »

Le marquage moléculaire (années 90) a ouvert la voie vers la localisation des locus impliqués dans la variation des caractères (QTL) et la calibration de modèles permettant de prédire la valeur génétique d'individus non évalués (Sélection Assistée par Marqueurs).

Sélection assistée par marqueurs
post QTL (SAM)

$$G = M = \sum_{QTL} \beta_{QTL} x_{QTL}$$



Sélection génomique
(SG)

$$G = M = \sum_m \beta_m x_m$$

Utilisation de tous les marqueurs

Whittaker et al, 2000. Remplacer l'estimation des effets des marqueurs par la méthode des moindres carrés par des approches pénalisées.
Rapidement testée par les sélectionneurs maïs au sein de population biparentales

Ex: Johnson, Plant Breeding Reviews, vol 24,
2003. Résultats obtenus dans une population
F2 avec 89 marqueurs

Model	Correlation coefficient		
	Grain moisture (%)	Yield (bu/acre)	Stalk lodging (%)
Least-squares	0.53	0.43	0.37
Generalized ridge	0.76	0.70	0.53

Principe développé ensuite par Meuwissen et al. 2001 (sur bovins laitiers)

« Sélection Génomique » ou « Sélection Génome-wide »

Dans années 90, grâce aux marqueurs moléculaires, premières évaluations de la diversité génétique et calculs de distances génétiques entre individus.

Bernardo, 1994 propose de prédire la valeur de nouveaux hybrides de maïs en utilisant leur apparentement avec des hybrides évalués en remplaçant l'apparentement sur le pédigrée par un apparentement moléculaire

BLUP modèle animal

$$y = \mu 1_n + Zu + e$$

$$u \sim N(0, A\sigma_u^2)$$

A basée sur le pédigrée



GBLUP

$$y = \mu 1_n + Zu + e$$

$$u \sim N(0, K\sigma_u^2)$$

$$K_{(i,i')} = \frac{\sum_{m=1}^M (x_{im} - 2f_m)(x_{i'm} - 2f_m)}{\sum_{m=1}^M 2f_m(1 - f_m)}$$

Où x_{im} génotype au marqueur de i au marqueur m (0, 1, 2), f_m la fréquence de l'allele de référence au marqueur m

2^{ème} approche de sélection génomique « GBLUP » équivalente à la ridge régression et popularisée en génétique animale dans les années 2000.

NB Bernardo, 1994 décompose u_{12} la valeur d'un hybride entre L1 et L2 en $AGC_1 + AGC_2 + ASC_{12}$

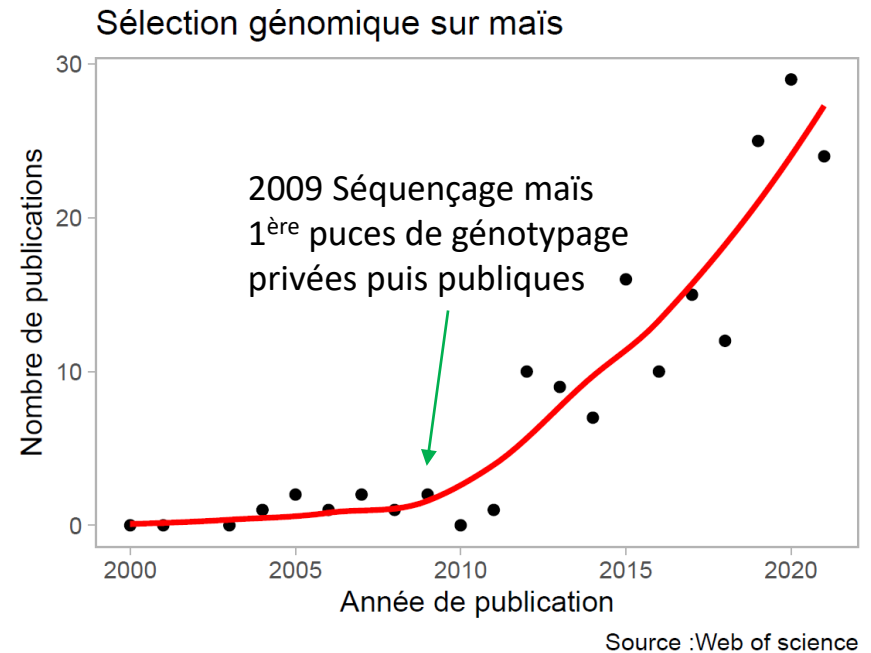
Les débuts de la sélection génomique chez le maïs

2009 séquençage du maïs et
apparition de puces de génotypage
dense à faible coûts

Actuellement 10-12 euros pour
plusieurs milliers de marqueurs
= coût d'une parcelle agronomique

Permet d'envisager la SG à l'échelle
d'un programme de sélection

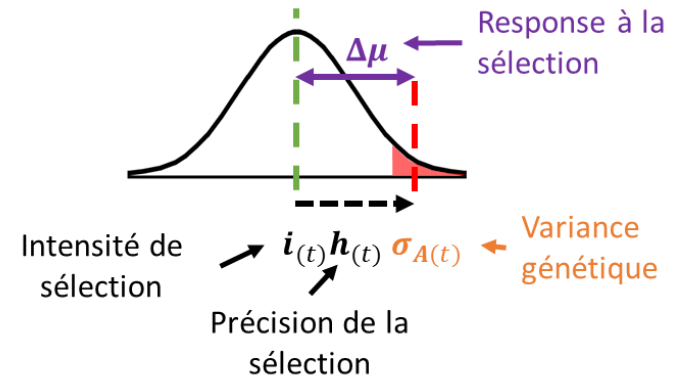
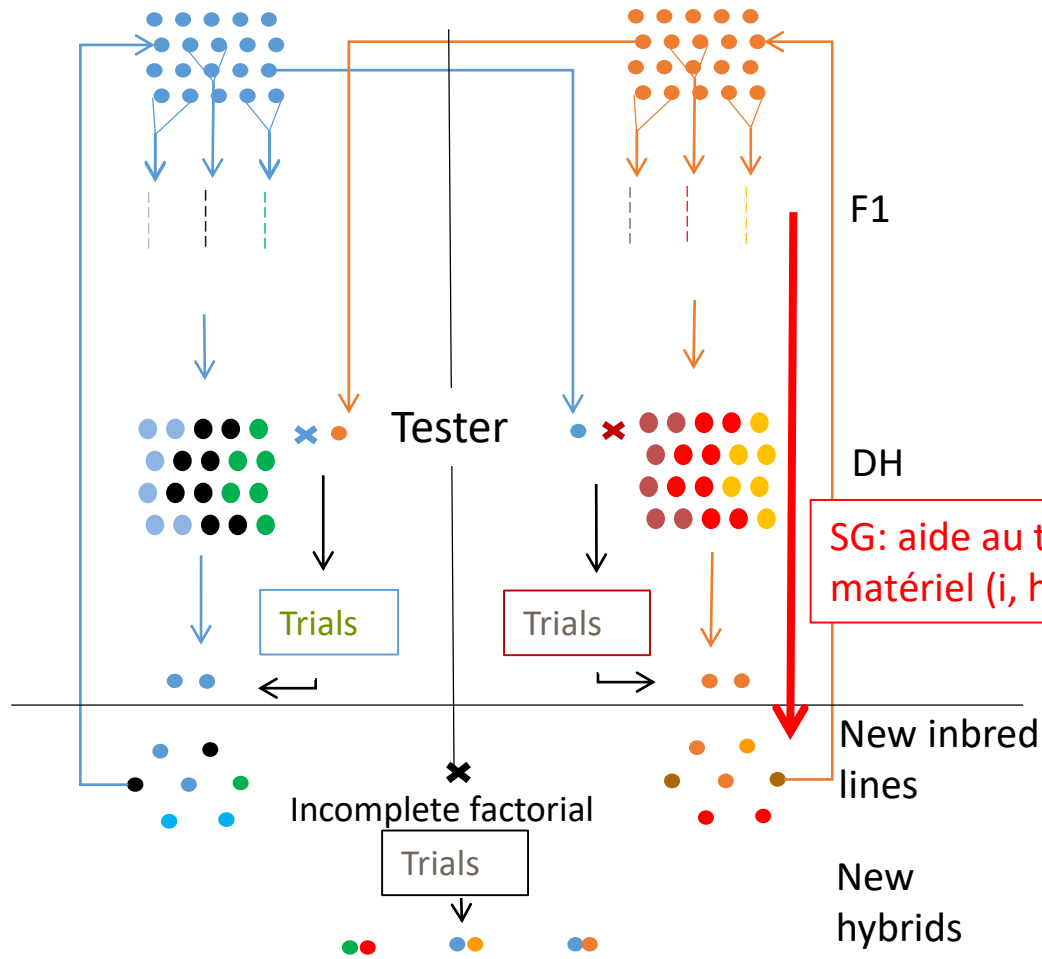
+/- 30 publications par an



Quel est l'intérêt de la SG chez le maïs ?

Intégration de la SG dans le schéma de sélection maïs

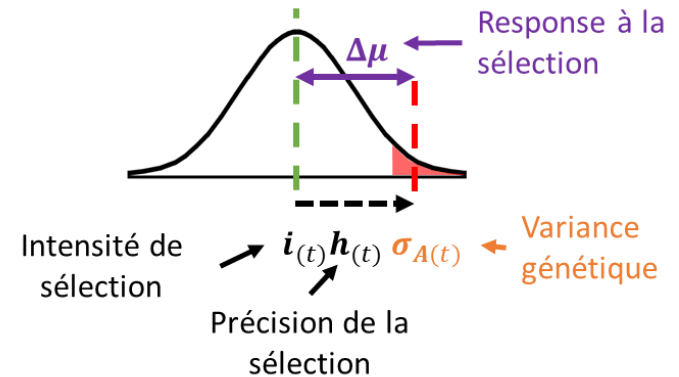
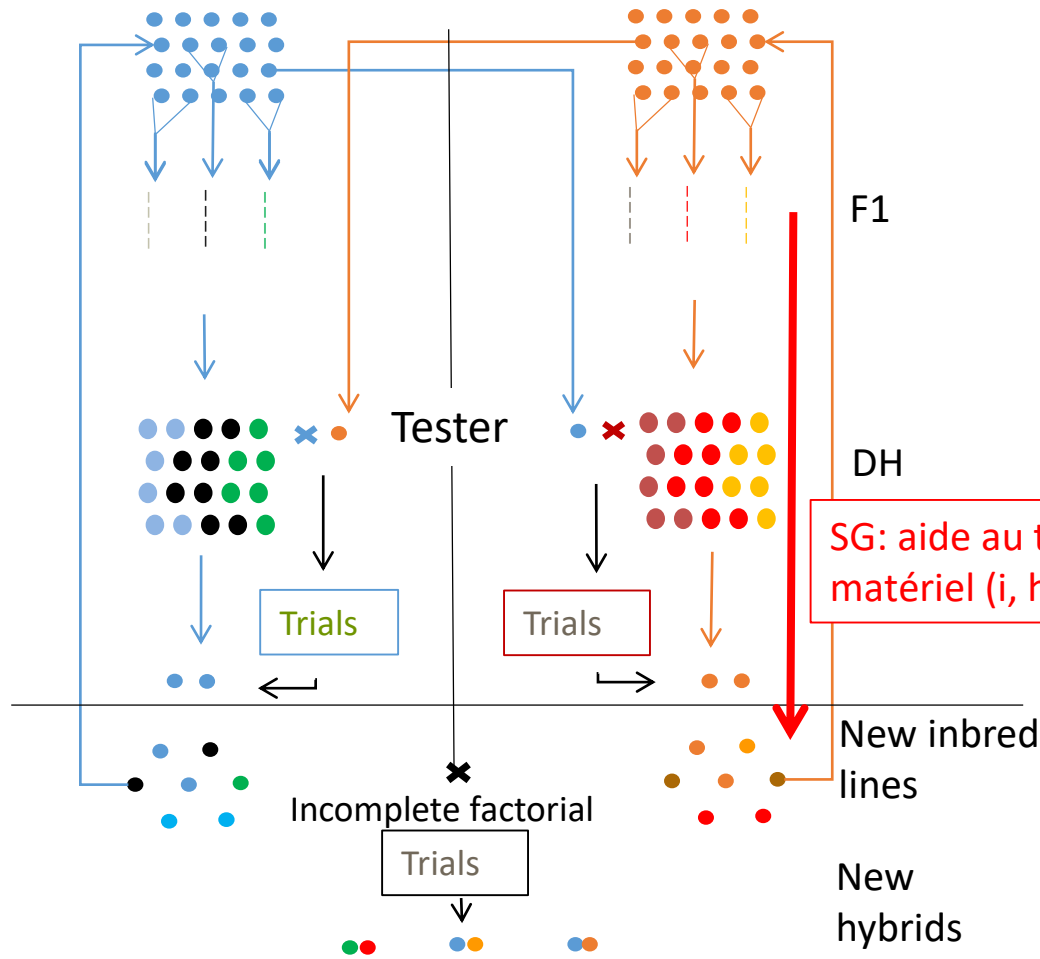
$$\Delta\mu / T = i_{(t)} h_{(t)} \sigma_{A(t)} / T$$



SG: aide au tri du matériel (i, h)

Intégration de la SG dans le schéma de sélection maïs

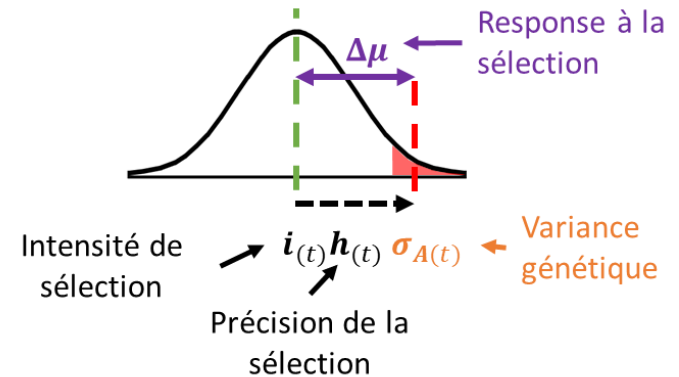
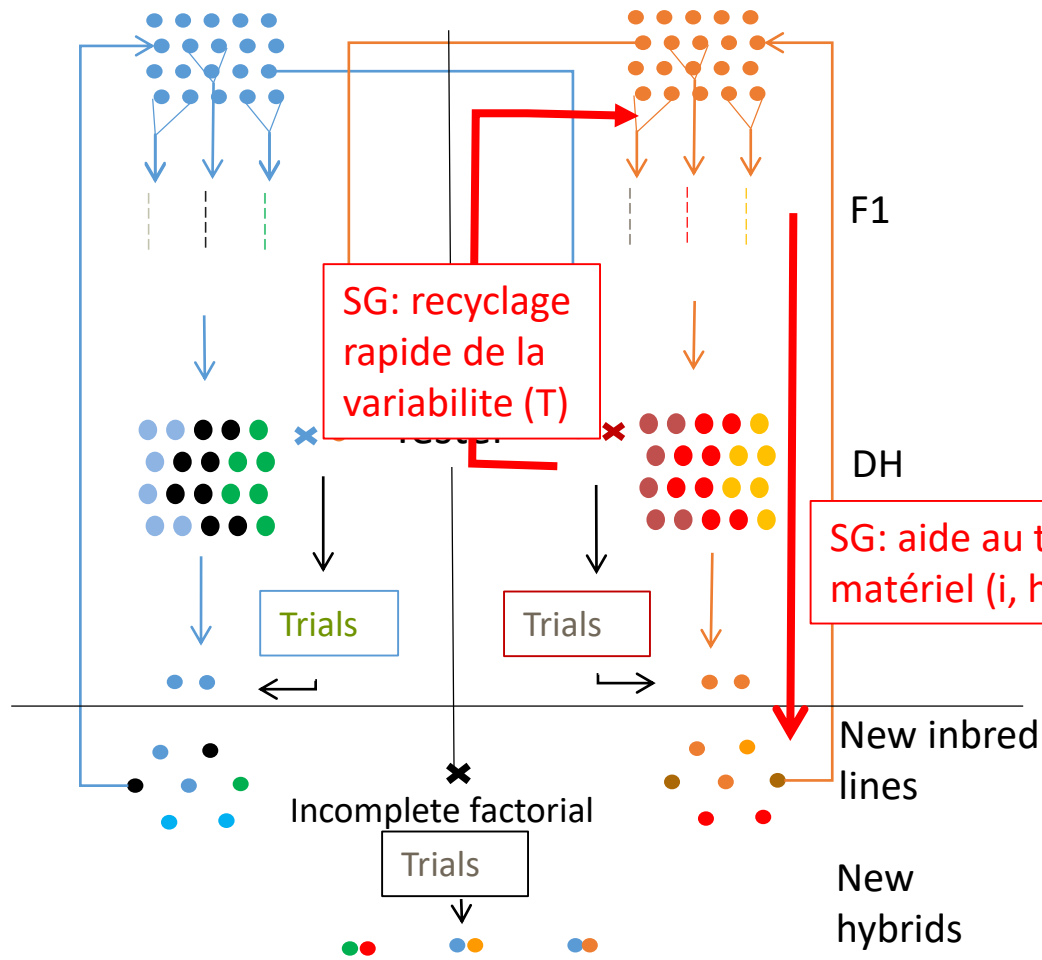
$$\Delta\mu / T = i_{(t)} h_{(t)} \sigma_{A(t)} / T$$



- Gestion plus globale des performances
- Remplacer (tout ou) une partie de l'évaluation des performances par des prédictions génomiques « Sparse testing » (ex: Atanda et al. TAG 2020..)

Intégration de la SG dans le schéma de sélection maïs

$$\Delta\mu / T = i_{(t)} h_{(t)} \sigma_{A(t)} / T$$



- Gestion plus globale des performances
- Remplacer (tout ou) une partie de l'évaluation des performances par des prédictions génomiques « Sparse testing » (ex: Atanda et al. TAG 2020..)

Quelle population utiliser pour calibrer les modèles ?

1ère option: Données historiques (rendement...) des cycles antérieurs (+ cycle en cours)
Option « facile » (même si GxE, diminution de l'apparement avec candidats...).

2ème option: Inclure de nouveaux caractères (caractères adaptatifs mesurés en plateformes en condition de stress...) non évalués en routine.



Ex Projet « Amaizing » : mesures en plateforme de paramètres de modèles écophysologiques en situation de déficit hydrique pour prédire GxE au champ.

Nombre de plantes évaluées est limité (<300)

Quels individus et quelles performances inclure dans les calibrations ?

Travaux pionniers sur maïs Rincent et al. Genetics 2012 cf synthèse récente Rio et al. in press

Prédiction des interactions GxE en combinant données de champs et de plateforme

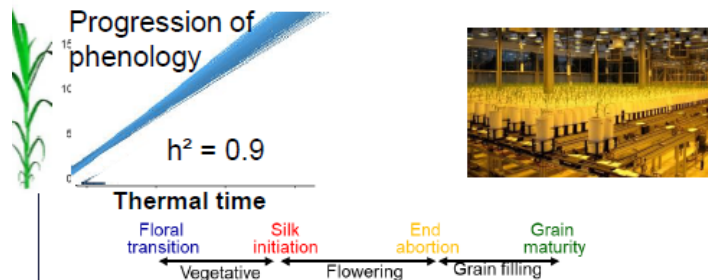
Genomic prediction of maize yield across European environmental scenarios

Millet et al 2019 Nature Genetics

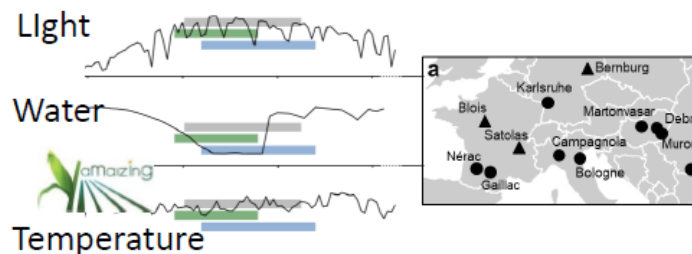


Ex: Données projets Drops/ Amaizing

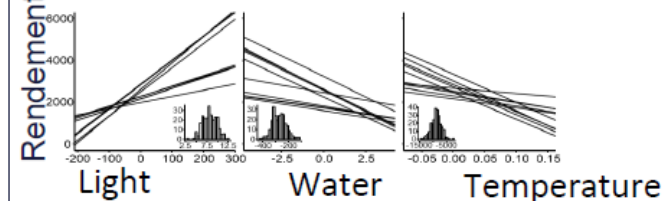
1. Platform, 250 geno



2. Field, Conditions felt by 250 genotypes in 25 fields



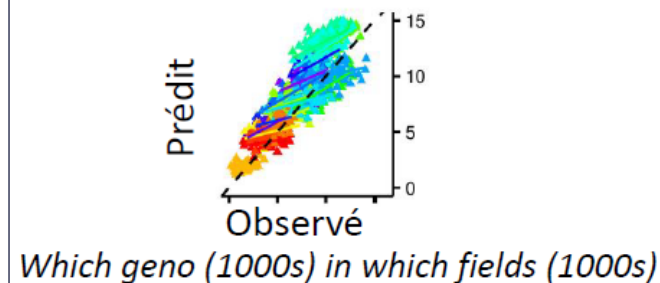
3. Response curves, 250 geno



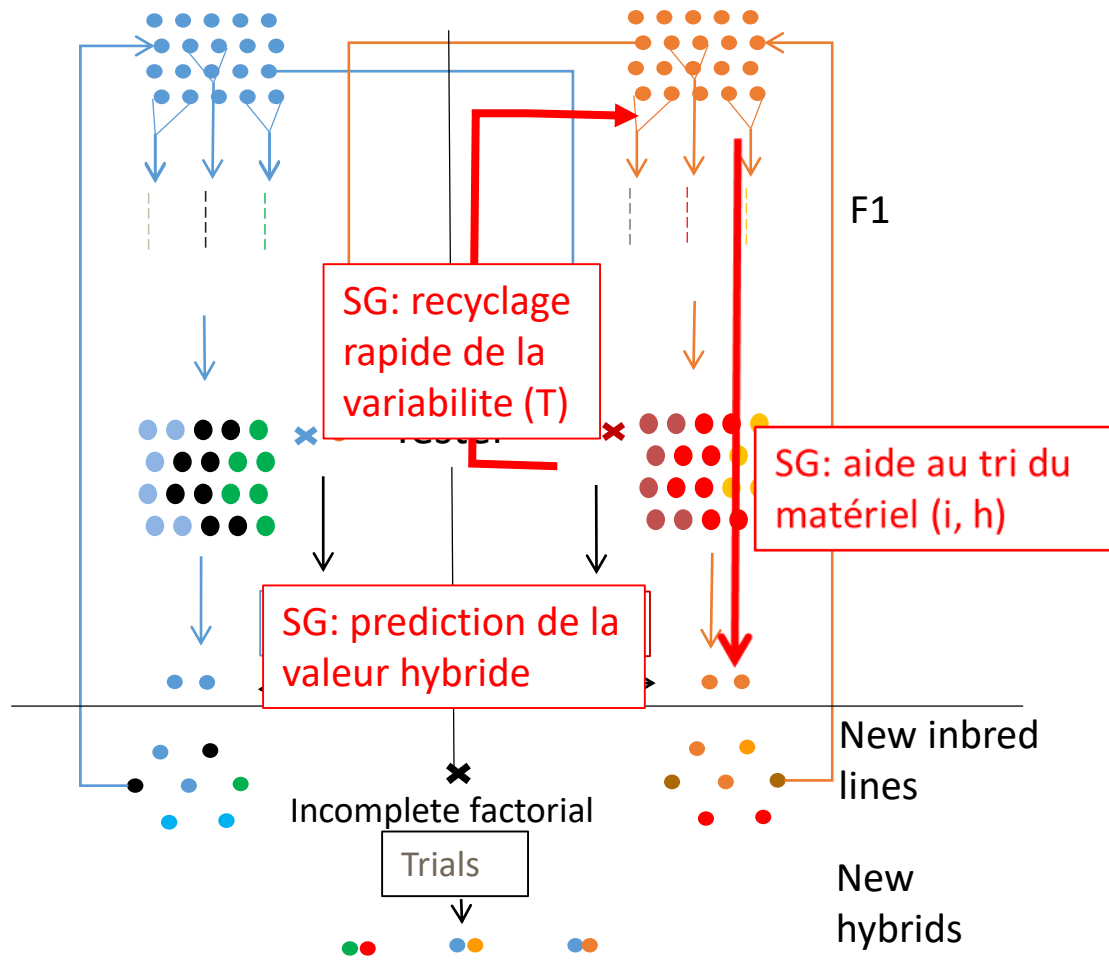
4. Genomic prediction, 250 geno

$$GY_{ij} = (\mu + e_j + g_i + \sum_{l=1}^s \beta_{l,i} cov_{l,i,j}) \times GW_i$$

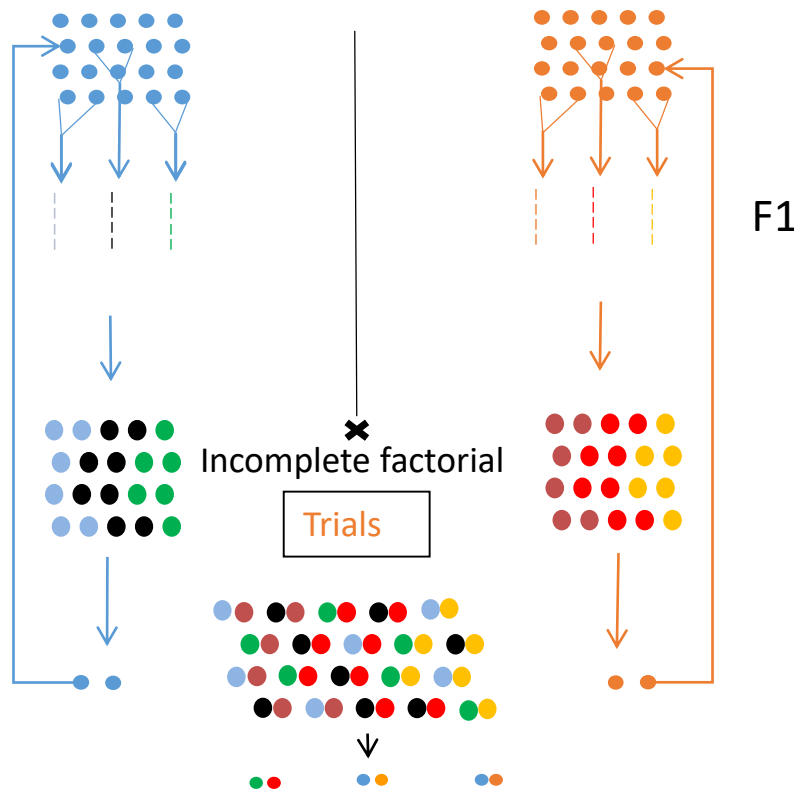
5. Test, New geno in new fields



Intégration de la SG dans le schéma de sélection maïs



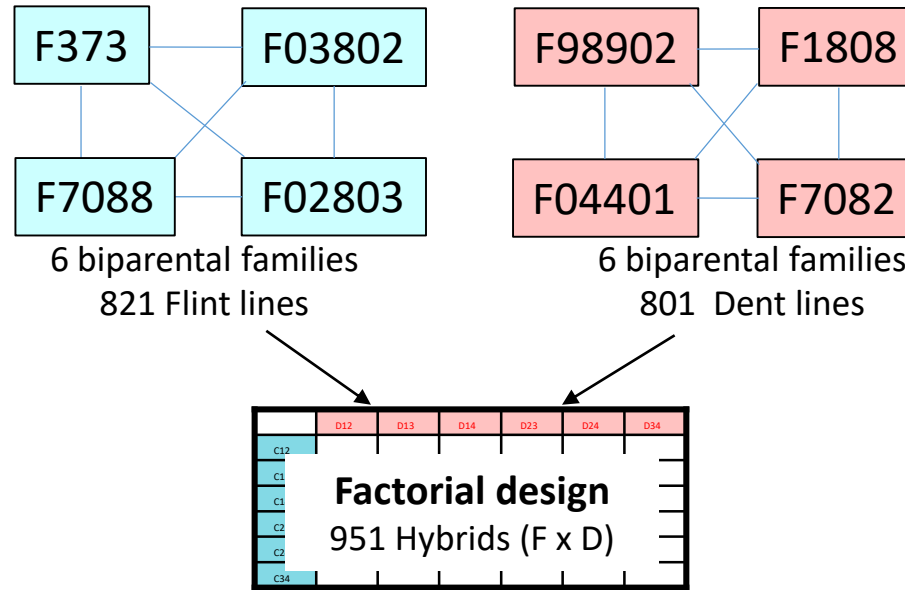
Intégration de la SG dans le schéma de sélection maïs



Au lieu d'évaluer les individus sur testeurs, puis d'identifier les meilleurs combinaisons hybrides, utilisation **d'un modèle de prédiction génomique calibré sur un factoriel incomplet** (un hybride par lignée) pour:

- Estimer les AGC des lignées
- Prédire les meilleures combinaisons hybrides (AGC + ASC)

=> L'utilisation d'un factoriel très incomplet (1 hybride / lignée) réduit par deux le nombre d'hybrides à évaluer par rapport à un dispositif sur testeur (un testeur/ groupe)



Th. H. Giraud



M2 A. Seye

Phase 1
8 essais
2013-14

Validation de la qualité de prédiction par cross-validation en utilisant le modèle de prédiction AGC + ASC proposé par Bernardo, 1994

NB: Adaptation récente du modèle de prédiction hybride notamment par Vitezica et al. 2017, Gonzalez-Dieguez et al. 2021.



Th. H. Giraud



M2 A. Seye

Résultats de cross-validation

	Rendement (tMS .ha ⁻¹)	Matiere sèche (%)	Date de floraison (j/1er Jan.)	UFL (UFL .kg ⁻¹)
Heritability (H ²)	0.814	0.892	0.890	0.740
% ASC / (AGC+ASC)	19%	14%	17%	8%
GBLUP				
r(AGCs, P)	0.644 sd 0.035	0.768 sd 0.024	0.700 sd 0.029	0.728 sd 0.027
r(AGCs + ASC, P)	0.652 sd 0.042	0.768 sd 0.024	0.697 sd 0.035	0.727 sd 0.027

- Prédictions génomiques efficaces (> prédictions sur QTL)
- Grâce aux marqueurs, bonne prediction de la valeur de lignées jamais évaluées au niveau hybride dans la population de calibration

Résultats similaires obtenus récemment dans la littérature (Kadam et al., 2016; Burdo et al. 2021....)



Th. H. Giraud



M2 A. Seye

Résultats de cross-validation

	Rendement (tMS .ha ⁻¹)	Matiere sèche (%)	Date de floraison (j/1er Jan.)	UFL (UFL .kg ⁻¹)
Heritability (H ²)	0.814	0.892	0.890	0.740
% ASC / (AGC+ASC)	19%	14%	17%	8%
GBLUP				
r(AGCs, P)	0.644 sd 0.035	0.768 sd 0.024	0.700 sd 0.029	0.728 sd 0.027
r(AGCs + ASC, P)	0.652 sd 0.042	0.768 sd 0.024	0.697 sd 0.035	0.727 sd 0.027

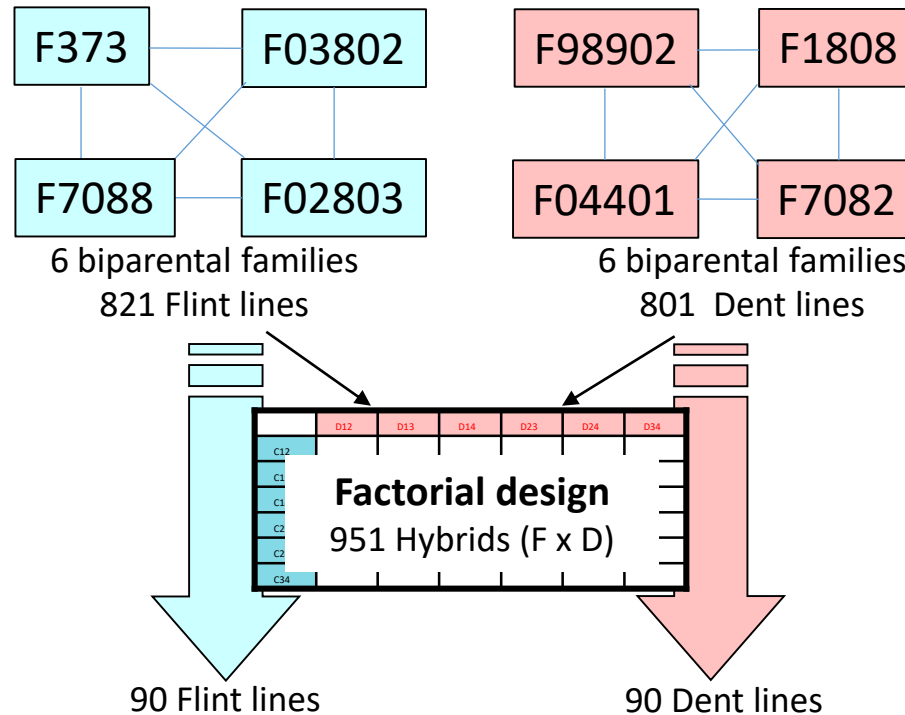
- Prédictions génomiques efficaces (> prédictions sur QTL)
- Grâce aux marqueurs, bonne prediction de la valeur de lignées jamais évaluées au niveau hybride dans la population de calibration

Est-ce qu'un dispositif factoriel est plus efficace qu'un dispositif sur testeur?

Schémas hybrides reconsidérés grâce à SG



M2 PhD A. Lorenzi



Phase 1
8 essais
2013-14

Factorial design

Dent testers

Flint testers

Phase 2
8 essais
2016-17

	D12	D13	D14	D23	D24	D34
C12						
C13						
C14						
C23						
C24						
C34						

360 Hybrids (F x D)

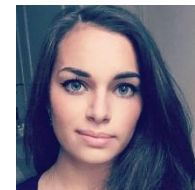
	T01	T02
C12		
C13		
C14		
C23		
C24		
C34		

2 Dent testers
180 Hybrids

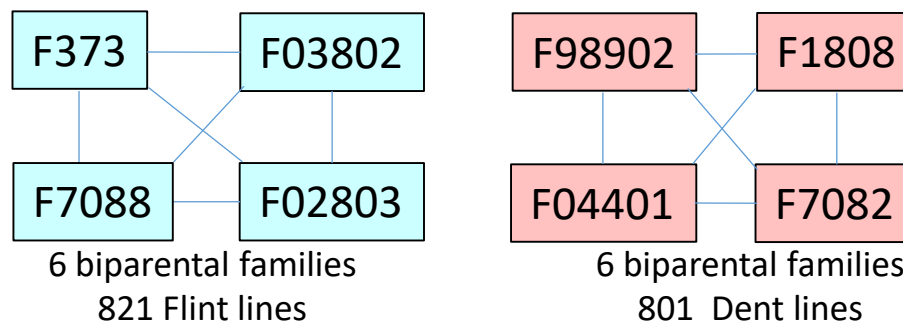
	D12	D13	D14	D23	D24	D34
TC1						
TC2						

2 Flint testers
180 Hybrids

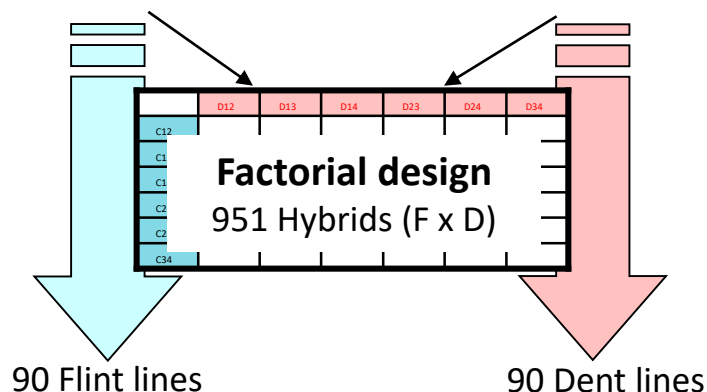
Schémas hybrides reconsidérés grâce à SG



M2 PhD A. Lorenzi



Phase 1
8 essais
2013-14



Phase 2
8 essais
2016-17

Factorial design

	D12	D13	D14	D23	D24	D34
C12						
C13						
C14						
C23						
C24						
C34						

360 Hybrids (F x D)

Dent testers

	T01	T02
C12		
C13		
C14		
C23		
C24		
C34		

2 Dent testers
180 Hybrids

Flint testers

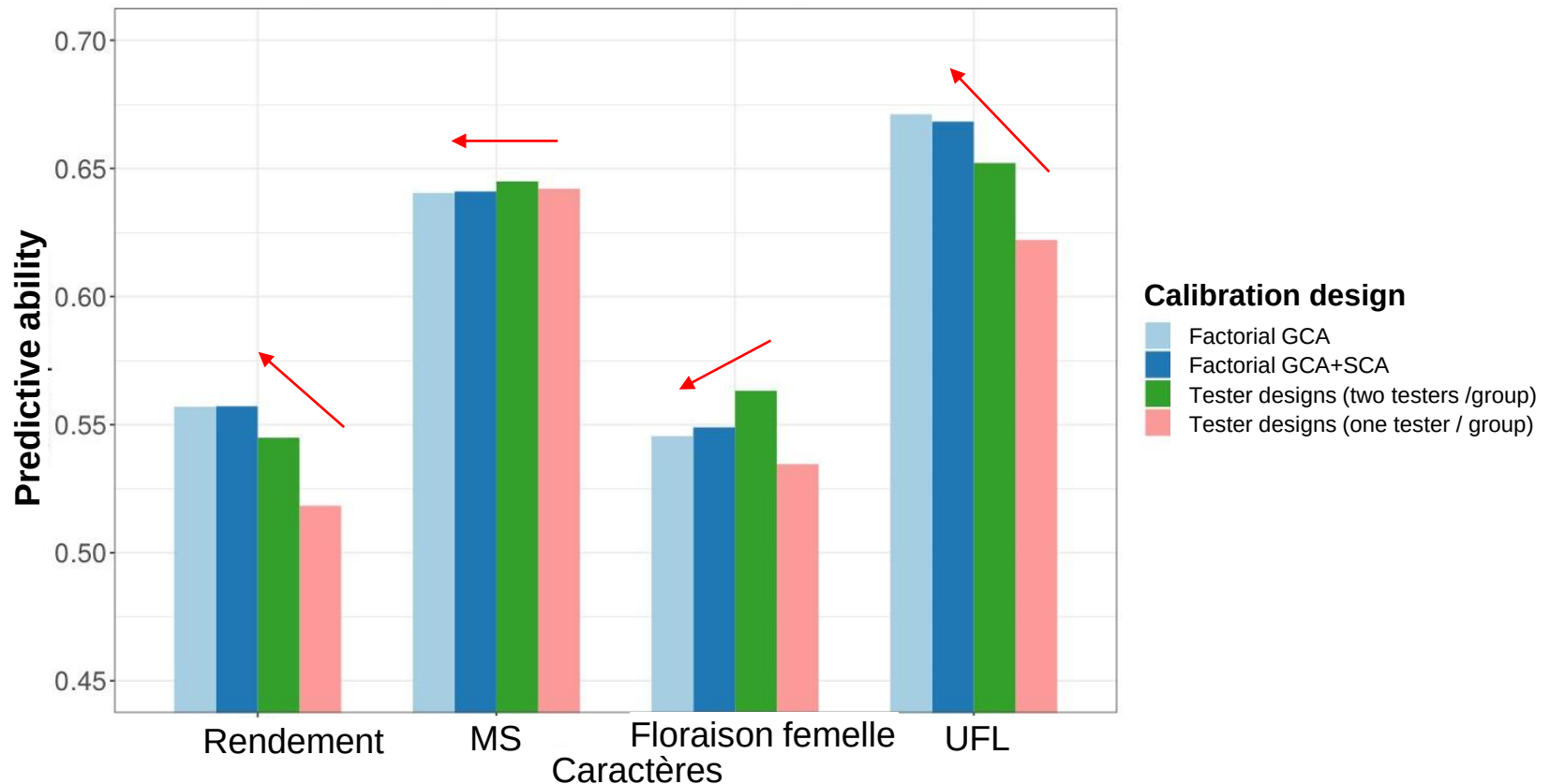
	D12	D13	D14	D23	D24	D34
TC1						
TC2						

2 Flint testers
180 Hybrids

Schémas hybrides reconsidérés grâce à SG

Lorenzi et al. soumis

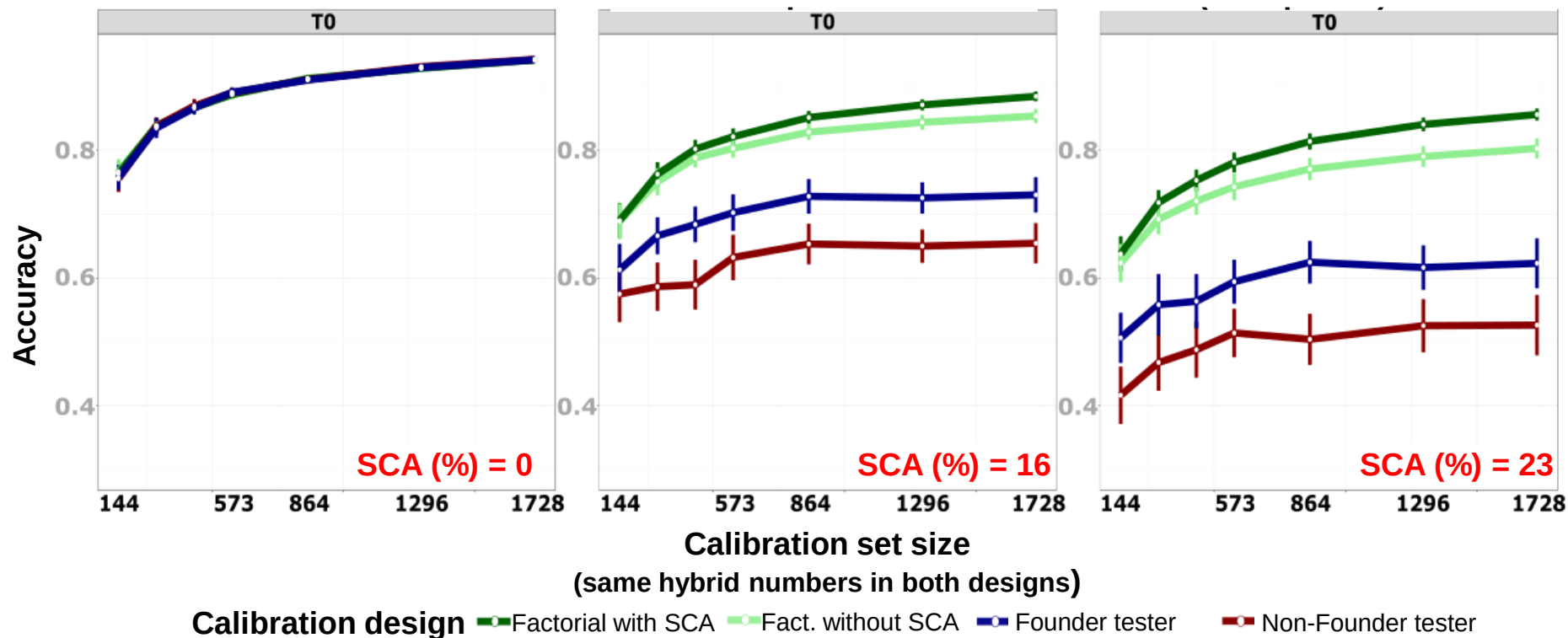
Comparaison calibrations sur factoriel –testeur pour prédire de nouveaux hybrides



- (Petit) avantage du dispositif factoriel sur le dispositif testeur à nb hybrides constant (pour rendement et UFL et en moyenne sur l'ensemble des caractères) même si les 2 testeurs = 2 des 4 lignées fondatrices des lignées du groupe complémentaire

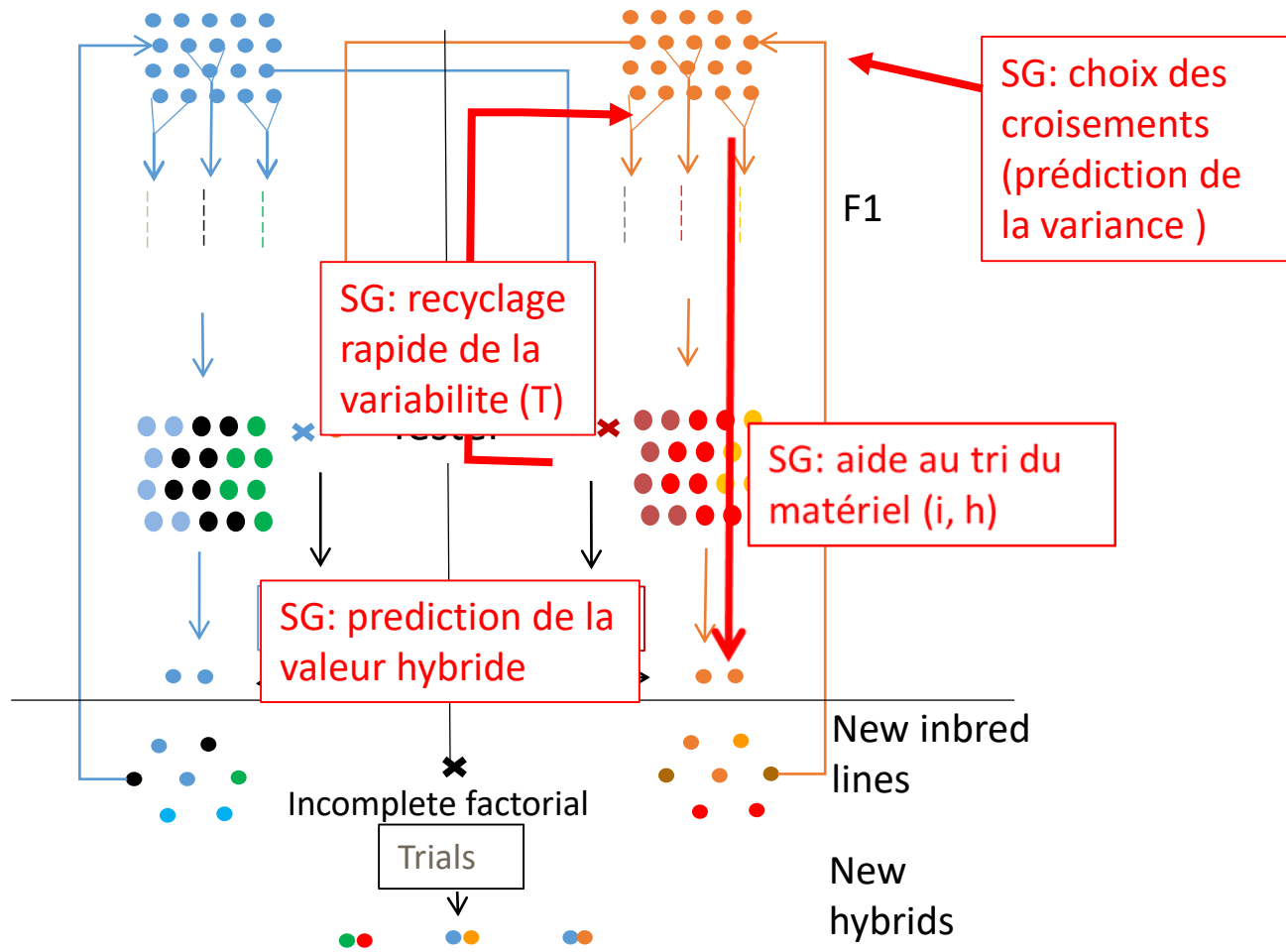
Comparaison des dispositifs testeur et factoriel par simulations

Seye et al. TAG 2020



➤ **Factoriel incomplet > testeur en présence d'ASC et précision plus stable**

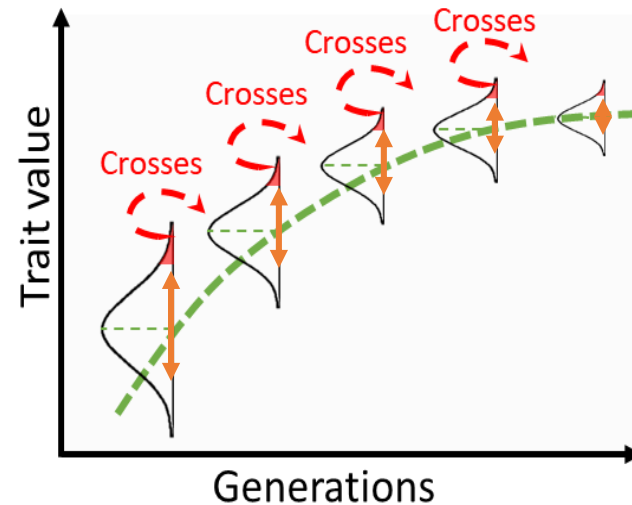
Intégration de la SG dans le schéma de sélection maïs



La variance génétique est essentielle en sélection

Equation du sélectionneur
(Lush 1937)

$$\Delta\mu = i_{(t)} h_{(t)} \sigma_{A(t)} / T$$

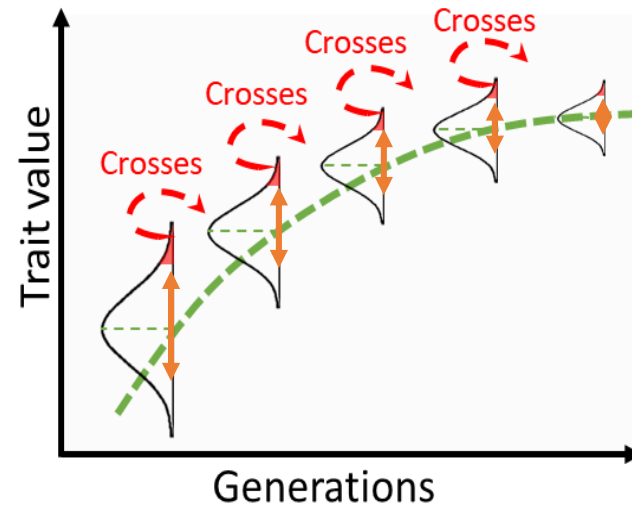


La sélection entraîne une perte de la diversité au cours des cycles
La SG peut conduire à une accélération de cette perte

La variance génétique est essentielle en sélection

Equation du sélectionneur
(Lush 1937)

$$\Delta\mu = i_{(t)} h_{(t)} \sigma_{A(t)} / T$$



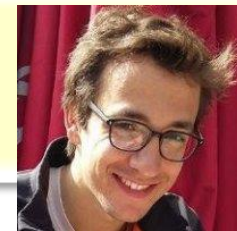
La sélection entraîne une perte de la diversité au cours des cycles
La SG peut conduire à une accélération de cette perte

Nécessité: -d'évaluer la diversité dans les programmes de sélection
-limiter sa perte tout en maintenant le progrès
-chercher de nouvelles sources de diversité

Utilisation de la SG pour mieux gérer la variabilité génétique dans les programmes de sélection

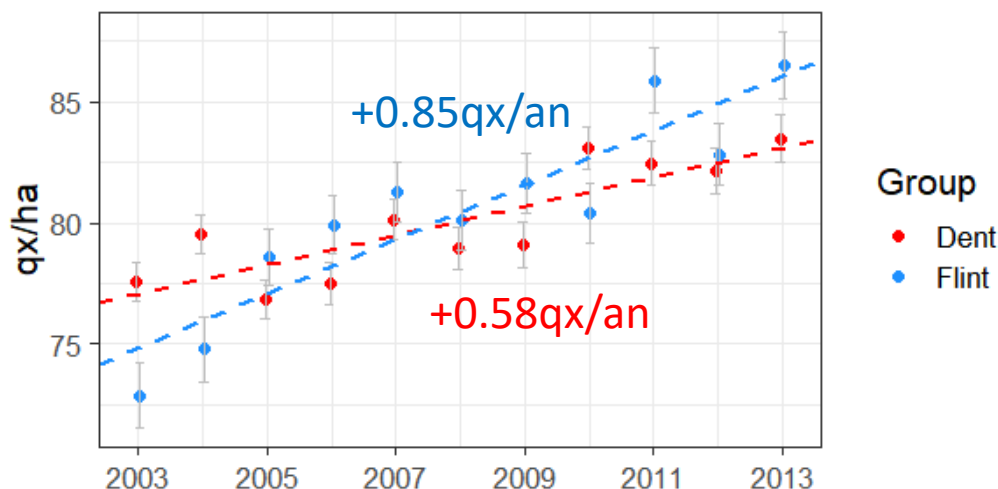


Allier et al. TAG, 2019a

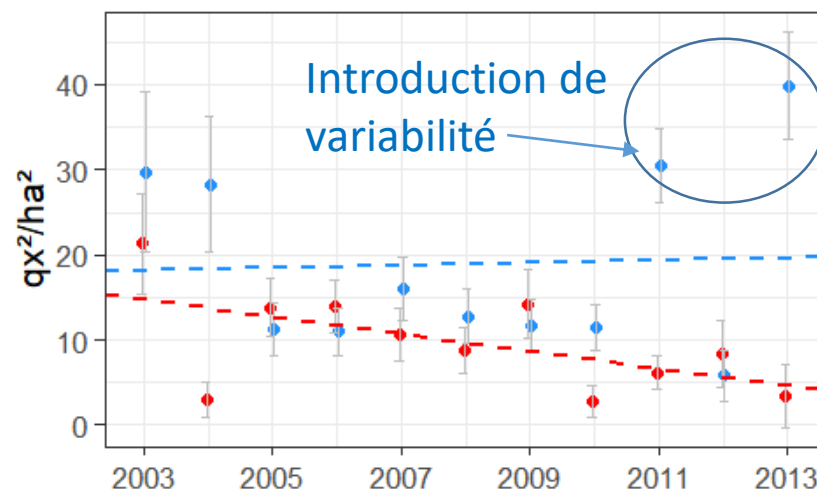


PhD of A. Allier
(cifre RAGT)

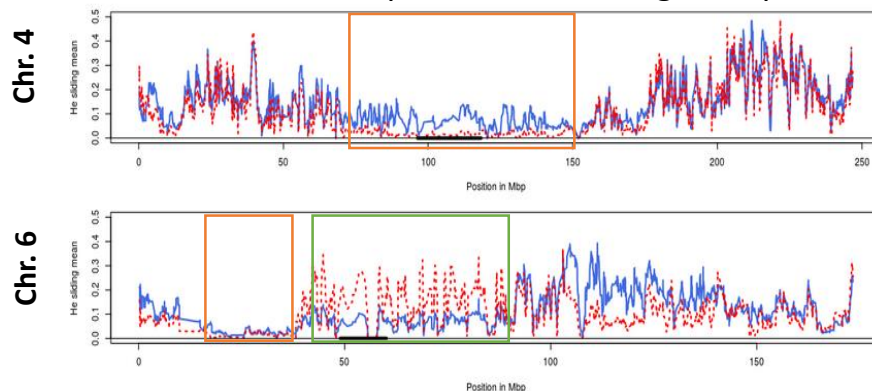
Evolution valeurs génétiques



Evolution de la variance



Evolution au cours du temps de la variabilité génétique le long du génome (Dentés)

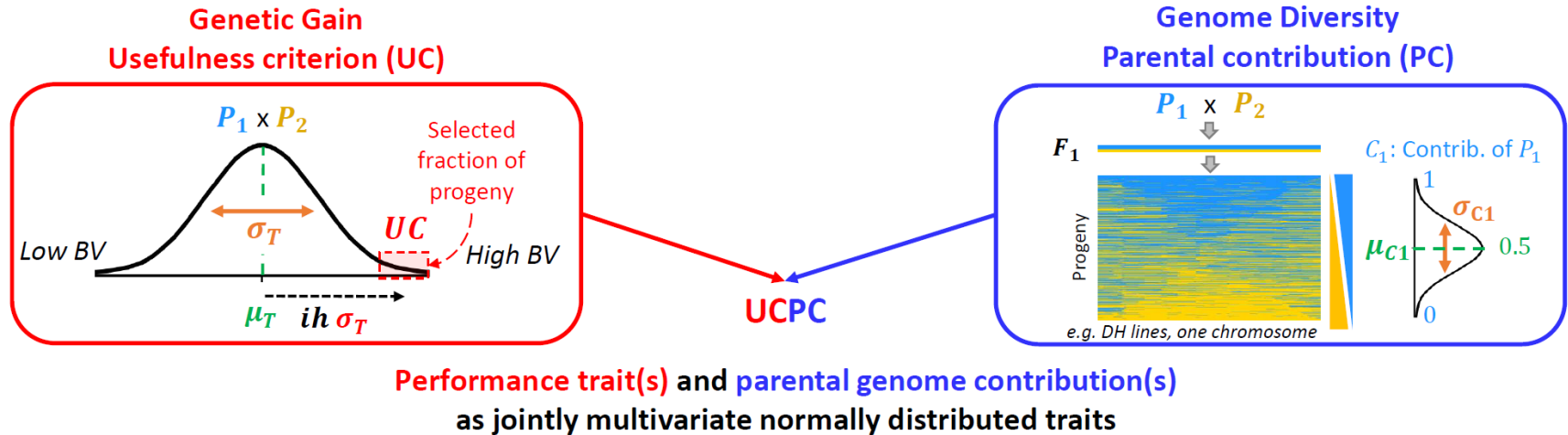


=> Comment utiliser la SG pour mieux gérer la variabilité en sélection?

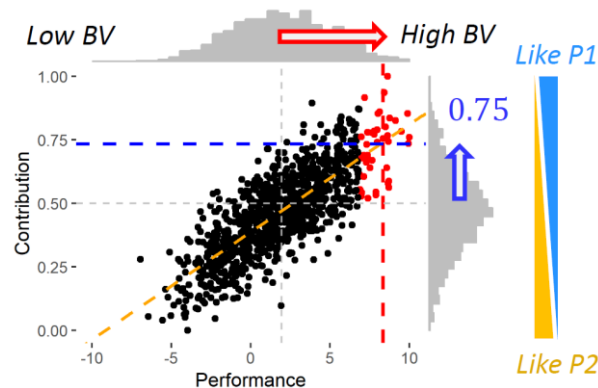
2003-2009
2010-2016

Utilisation de la SG pour mieux gérer la variabilité génétique dans les programmes de sélection

Allier et al. G3, 2019



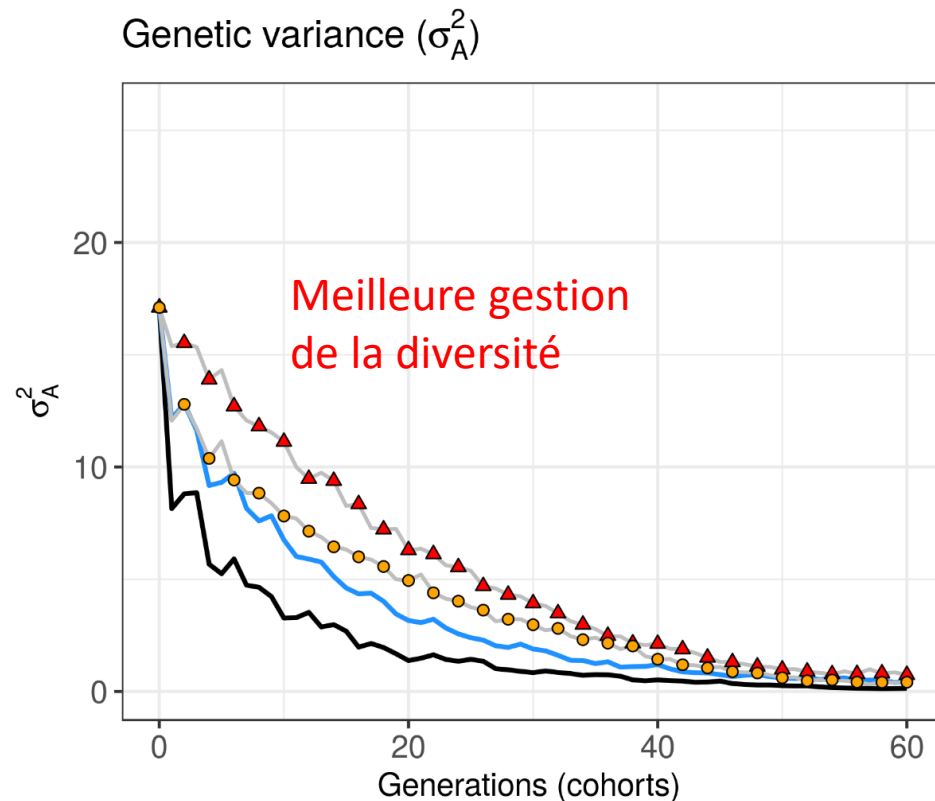
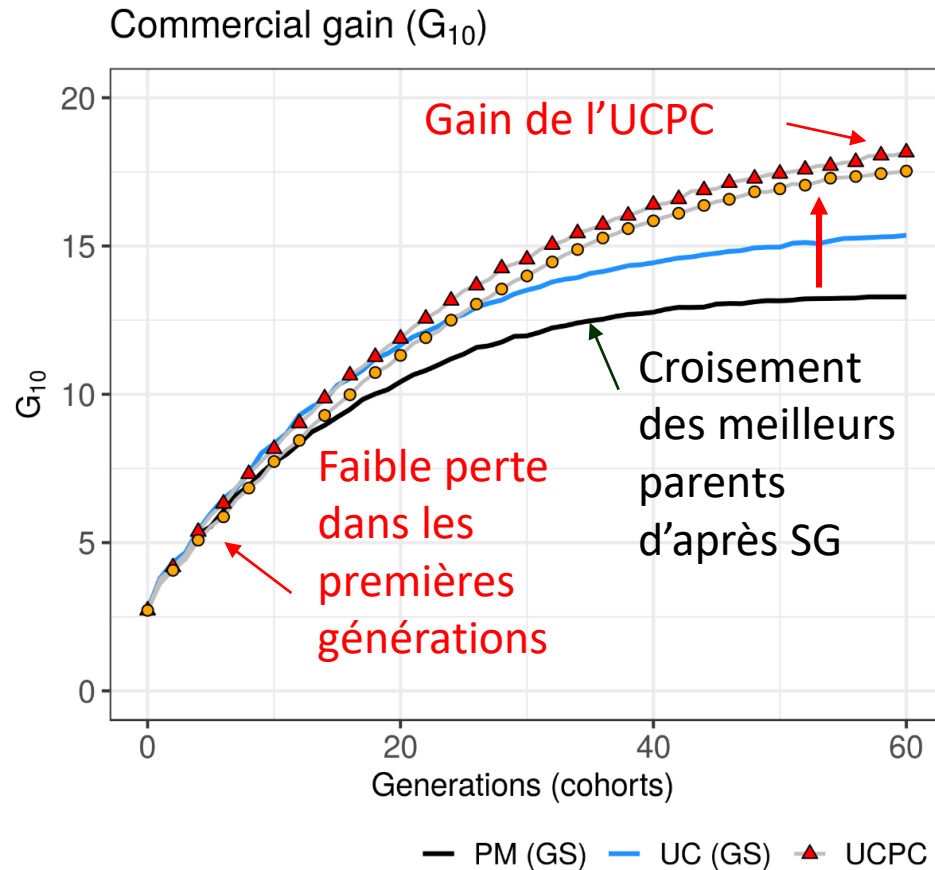
$$\text{If } P_1 > P_2 \Rightarrow \sigma_{T,C1} > 0$$



Application: choix des meilleurs croisements à chaque cycle de sélection pour maximizer le progrès et limiter la perte de diversité

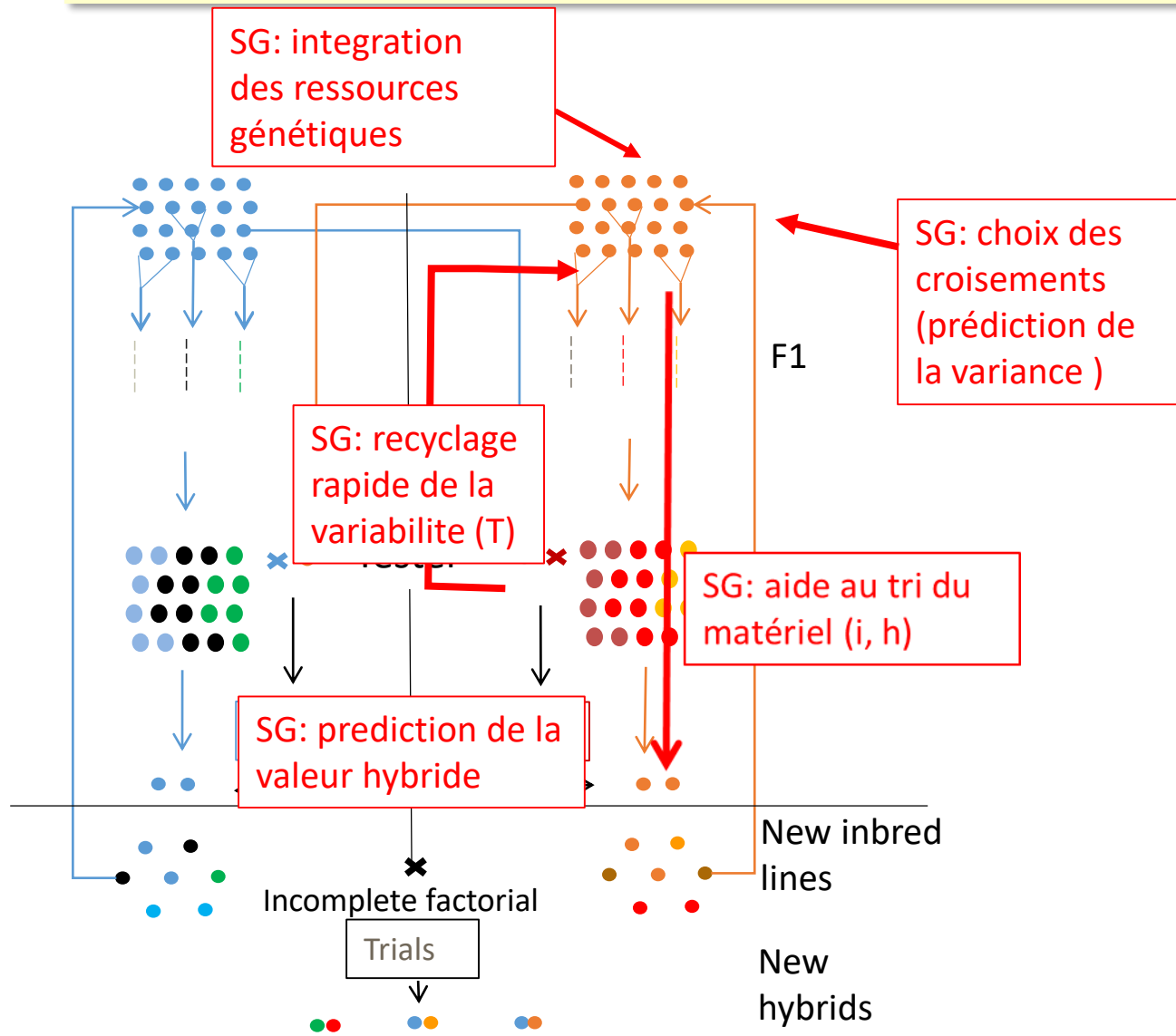
Utilisation de la SG pour mieux gérer la variabilité génétique dans les programmes de sélection

Allier et al. Front. Genet 2019



=> La prise en compte de la complémentarité entre lignées lors du choix des croisements permet une meilleure utilisation de la variabilité génétique

Intégration de la SG dans le schéma de sélection maïs



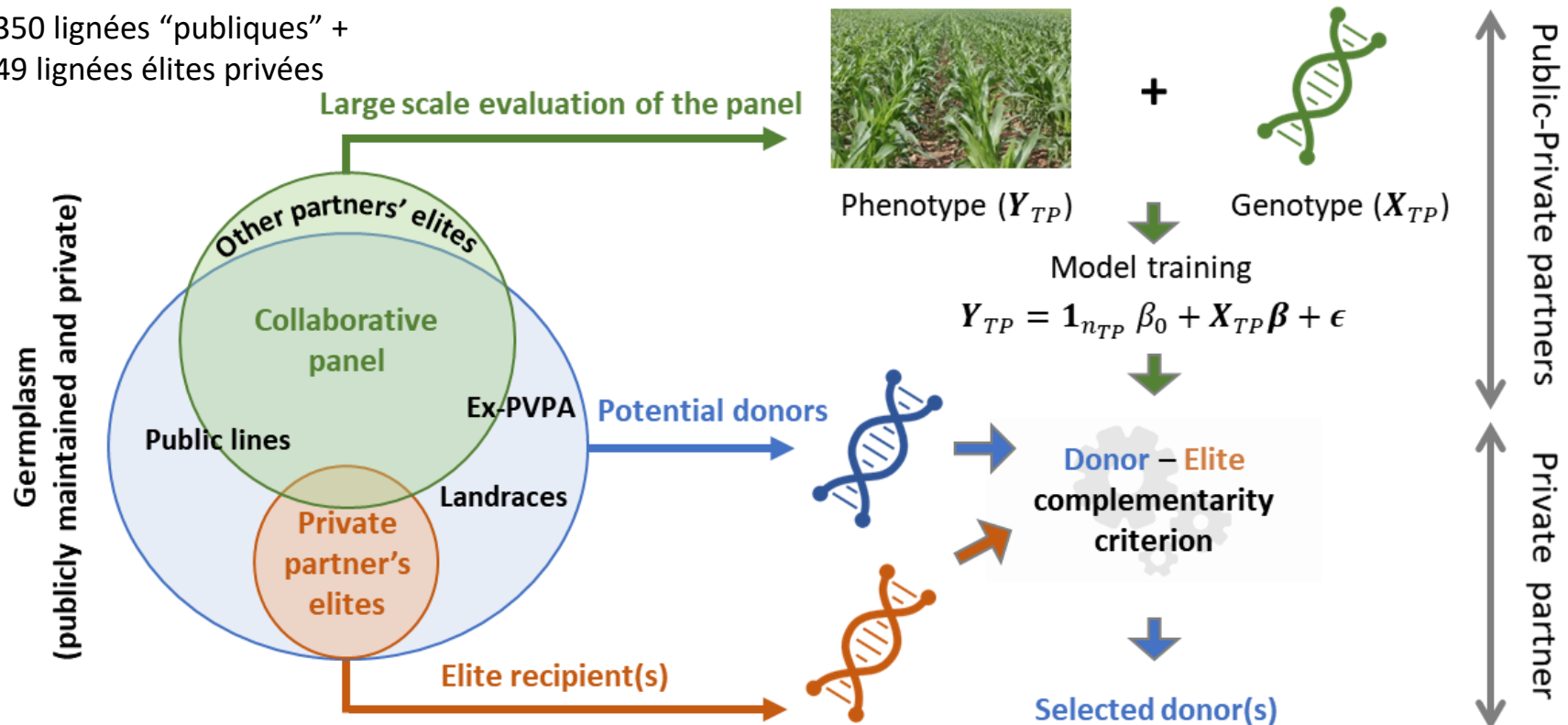
Introduction de nouvelles sources de variabilité génétique dans les programmes de sélection

Allier et al., TAG 2019b

Stratégie possible : utilisation de panels collaboratifs ex panel “Amaizing”



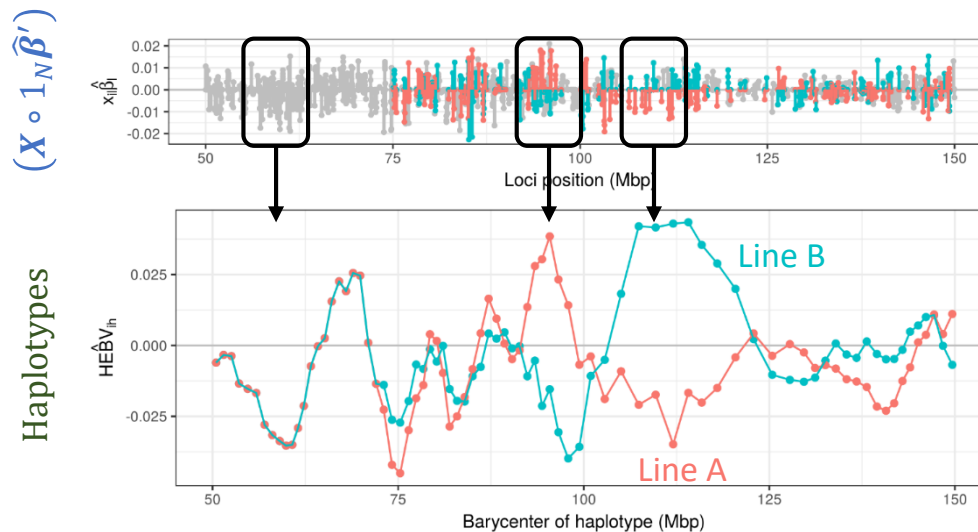
350 lignées “publiques” +
49 lignées élites privées



Utilisation des prédictions génomiques pour identifier des donneurs d'intérêt dans un panel

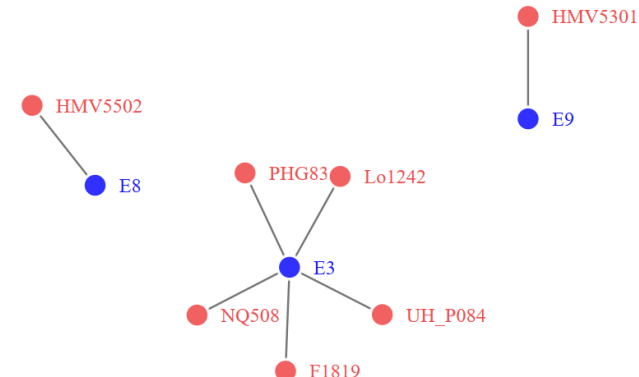
Allier et al. TAG 2019b

Utilisation des effets des marqueurs ($\hat{\beta}$) pour définir des blocs haplotypiques + / -



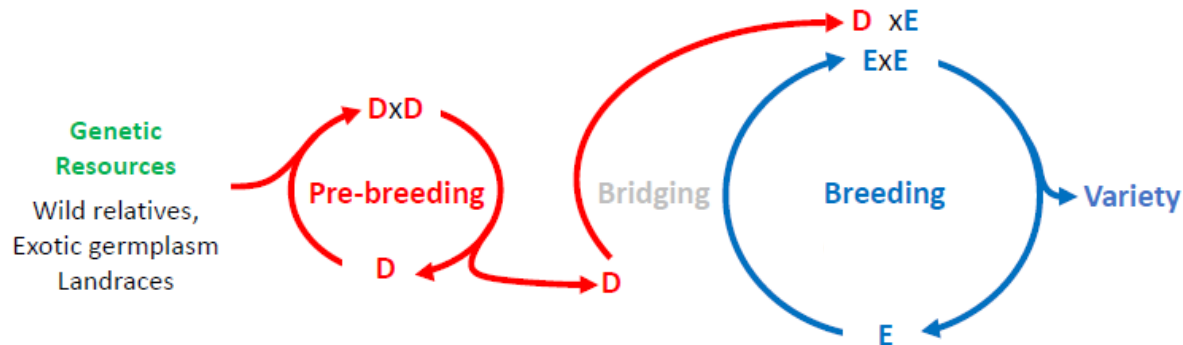
- Effets des allèles de Lignée A
- Effets des allèles de Lignée B
- Effets similaires A et B

Processus itératif pour choisir les lignées complémentaires du pool de lignées élites et choix des croisements

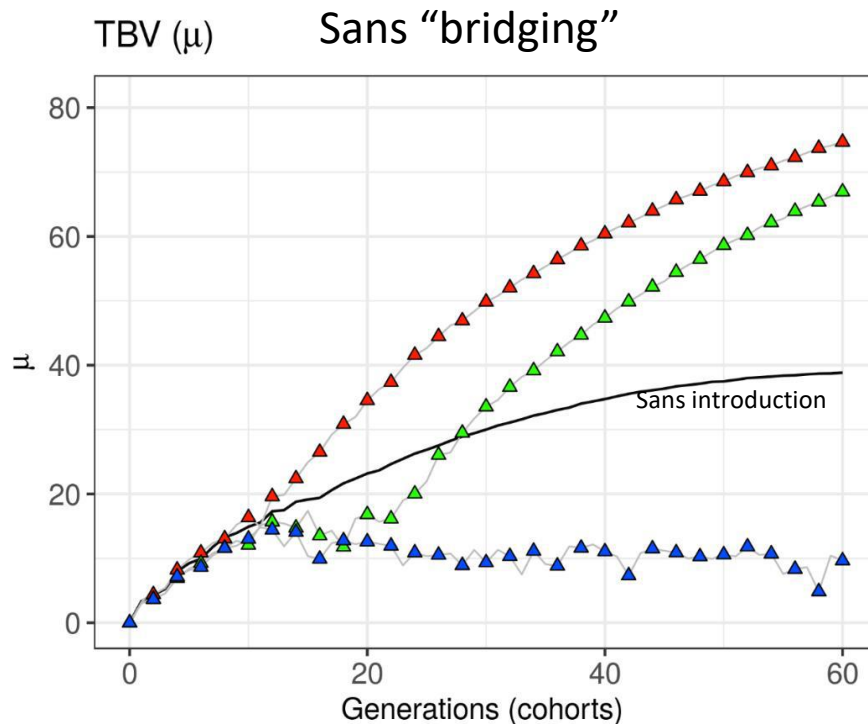


Intérêt d'introduire de la diversité dans les programmes de sélection

Allier et al., 2020



D = Donor
DE = Bridging progeny
E = Elite progeny

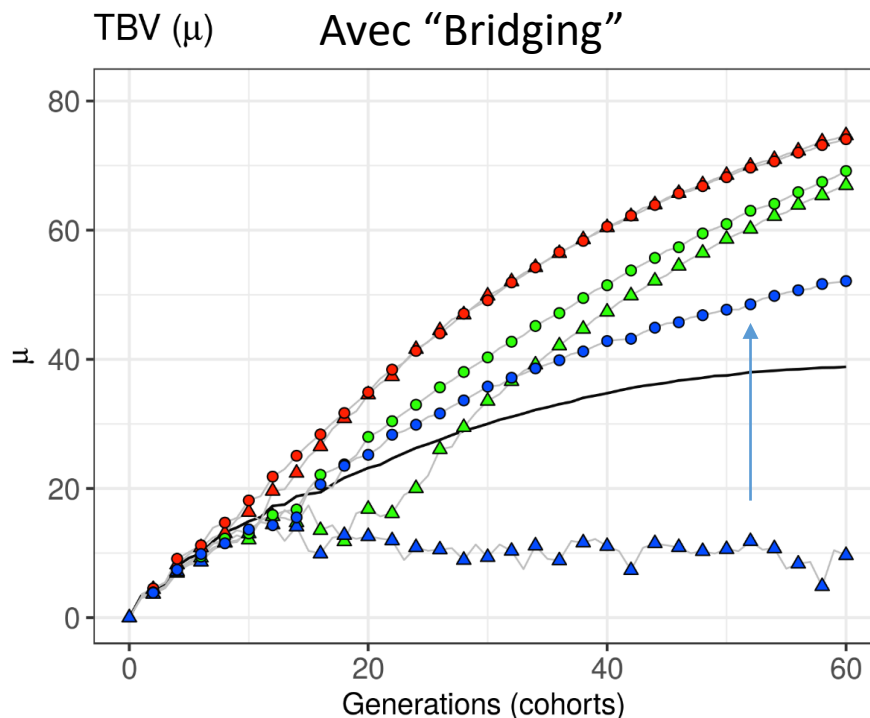
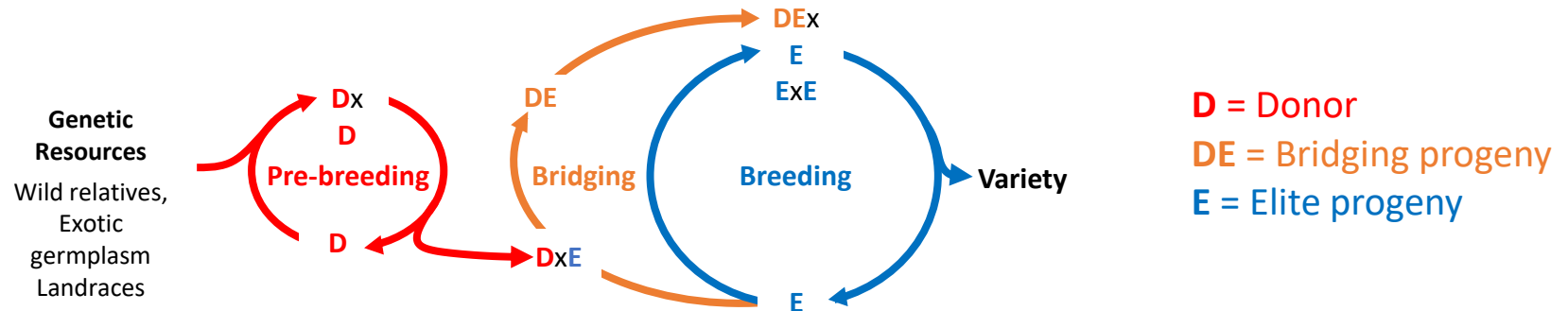


- **Donneurs récents (5 ans)**
- **Donneurs améliorés mais anciens (“ex pvp” 20 ans)**
- **Donneurs issus du panel (20 à 80 ans)**

— Benchmark ▲ Nobridging_5y ▲ Nobridging_20y ▲ Nobridging_Panel

Intérêt d'introduire de la diversité dans les programmes de sélection

Allier et al., BMC 2020



- **Donneurs récent (5 ans)**
- **Donneurs améliorés mais anciens ("ex pvp" 20 ans)**
- **Donneurs issus du panel (20 à 80 ans)**

— Benchmark ▲ Nobridging_5y ▲ Nobridging_20y ▲ Nobridging_Panel
 ● Bridging_5y ● Bridging_20y ● Bridging_Panel



Conclusions

- Contrairement à la SAM, fort intérêt de la SG chez le maïs (nombreuses publications public-privé sur le sujet)
- Essor grâce à la réduction du coût du marquage, mais aussi au développement des outils numériques (bases de données, accroissement des capacités de calculs...) et des développements statistiques.
- Nombreux avantages à différentes étapes pour
 - ✓ réduire le phénotypage,
 - ✓ accélérer les cycles
 - ✓ revisiter les schémas hybrides
 - ✓ gérer la variabilité génétique...
 - ✓ Introduire de la diversité : projet en cours Promaïs “ValRG”



Perspectives

- Utiliser de nouvelles données « omiques »?(transcriptomique, protéomique, métabolomique, réseaux de gènes, couplage avec modèles écophysiologiques pour prédire GxE, phénomique)
 - Identification de locus « clés » pondérés différemment dans les modèles de prédictions génomiques
 - Utilisation de prédicteurs au delà du génotype SNP (phénomique= génotype exprimé) ex données NIRS (Rincen et al. 2018)

Travaux en cours sur des données générées dans Amaizing
- Utiliser la prédiction génomique pour identifier de nouvelles sources de diversité pour de nouveaux caractères?
 - SG pour prédire la performance de populations et identifier les plus intéressantes?
 - Quelle stratégie de SG pour optimiser les introgressions dans le matériel élite?
- Vers une sélection 4.0 (Ramstein et al. 2019)? Combiner SG et l'édition de génome ? Peu de gènes majeurs identifiés (phénologie, domestication) et difficultés techniques pour l'édition du maïs.
 - Valider de locus candidats, des facteurs de régulation... avant SG
 - Adapter rapidement des ressources génétiques exotiques
 - Néodomestiquer le maïs à partir de la téosinte?

Remerciements

L'équipe GQMS

A. Charcosset, S. Nicolas
J. Fievet, T. Mary-Huard, R. Rincenc, C. Bauland
D. Madur, V. Combes, S. Pin

Etudiants:

H. Giraud, A. Seye, A. Lorenzi, S. Rio, A. Allier

Collaborateurs:

MIA: B. Mangin, S. Robin

GA: D. Laloé, Z. Vitezica

EA: F. Tardieu (LEPSE)

UE SMH: C. Palaffre

Partenaires des projets Amaizing,
Promaïs (SAMMCR, ValRG), Cornfed et
Selgen R2D2

RAGT2n:

C. Lehermeier
S. Teyssède



Partenaires privés :

Lidea, Syngenta, Biogemma, Limagrain,
KWS, RAGT, Maseeds, Corteva, Arvalis
Partenaires de Promaïs, Cornfed, Amaizing



