

# LA SELECTION VEGETALE DANS L'ERE NUMERIQUE

---

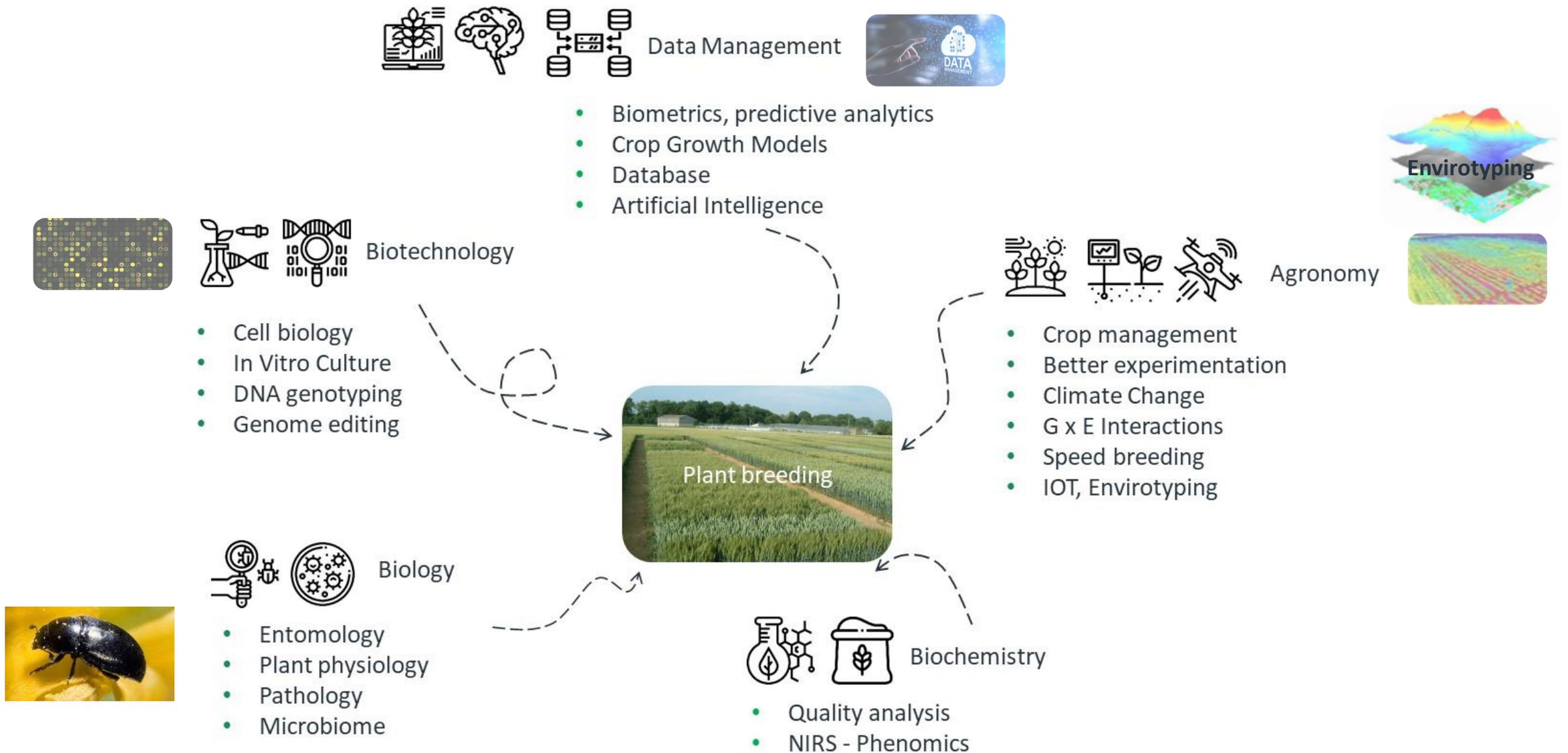
PHILIPPE DUFOUR  
RAGT2N

---

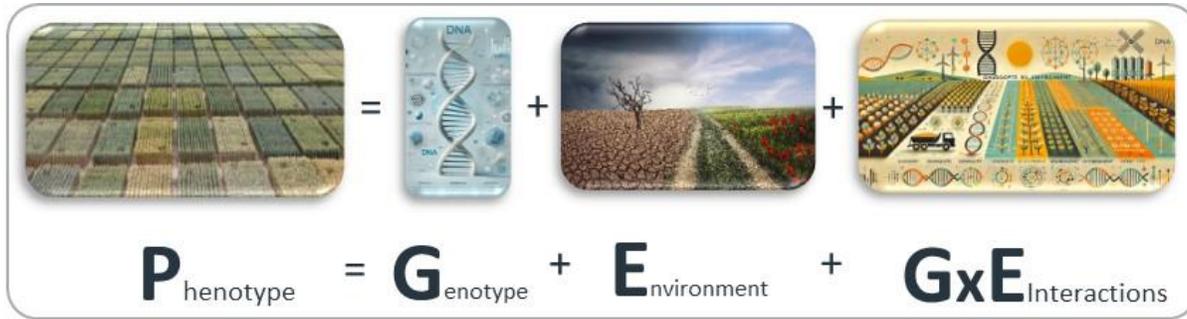


think  
**SOLUTIONS**  
think **RAGT**

# Les sélectionneurs utilisent des informations de plus en plus variées et complexes

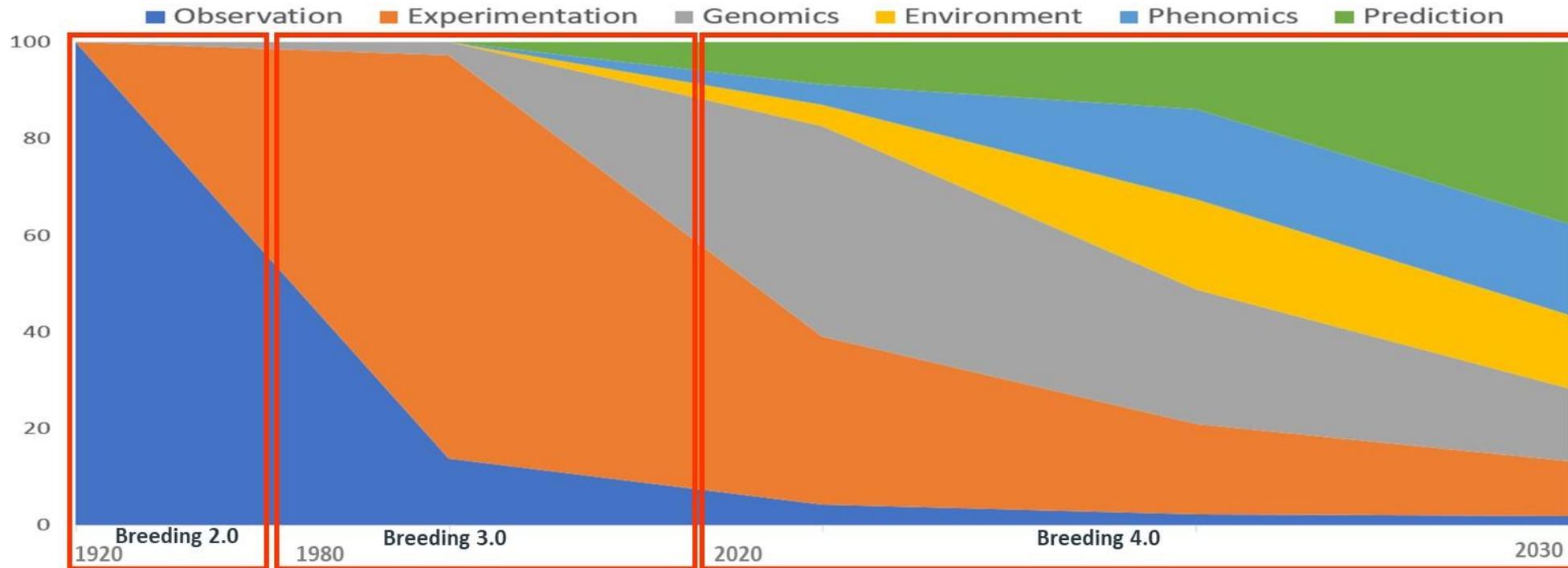


# De la sélection phénotypique à la sélection prédictive

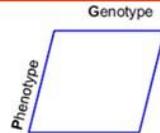


changements majeurs :

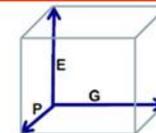
- Génotypage à haut débit
- Puissance de calcul / AI



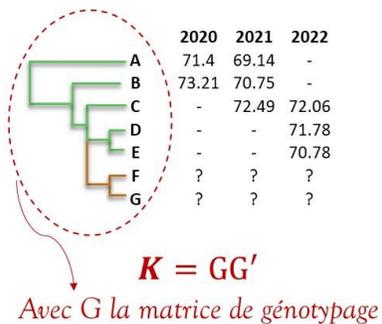
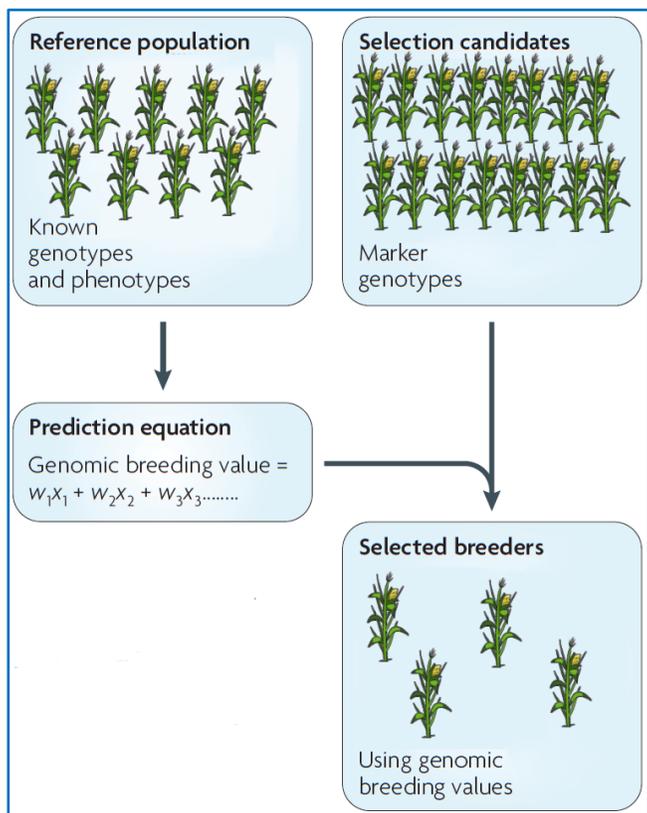
Phenotype



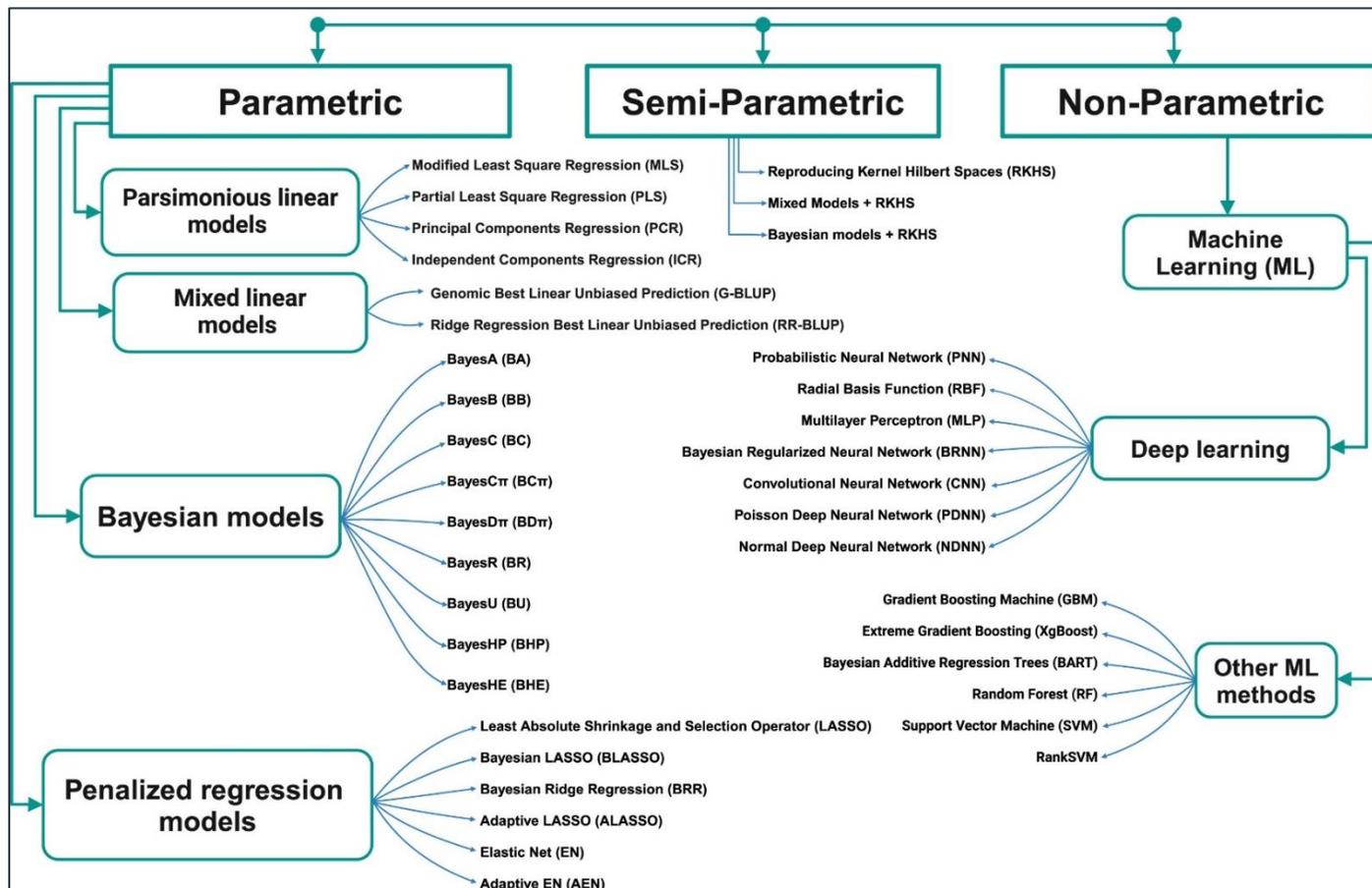
PREDICTION



Aperçu schématique de la construction du modèle GS (source Steven Maenhout)



Liste des modèles statistiques d'apprentissage automatique actuellement utilisés pour la prédiction génomique

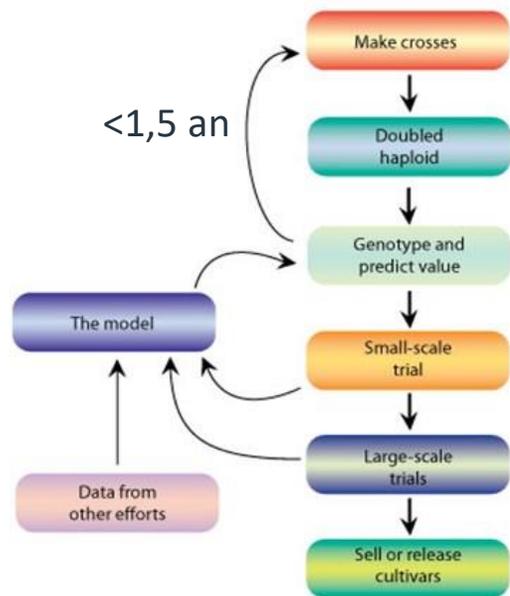
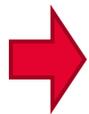
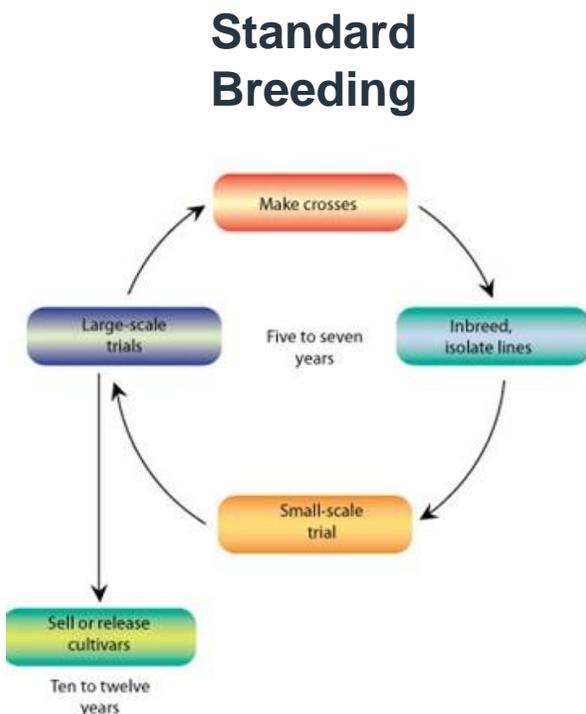


Le deep learning est particulièrement efficace dans les **tâches de prédiction**, où la quantité de données est énorme et où les relations entre les variables sont complexes et non linéaires, par exemple pour la reconnaissance d'image, l'estimation des effets GxE ....

# Des nouvelles méthodes de sélection pour accélérer le gain génétique



## Sélection génomique



10 ans

< 6 ans

Augmenter le gain génétique à différentes étapes de la sélection en combinant efficacement les données phénotypiques et génomiques

- Gain en précision, en temps et pour une meilleure estimation des effets génétiques
- Gestion dynamique et maîtrisée de la diversité génétique
- Nouvelles perspectives dans les plans de croisement et la constitution de variétés hybrides sur la base de modèles prédictifs

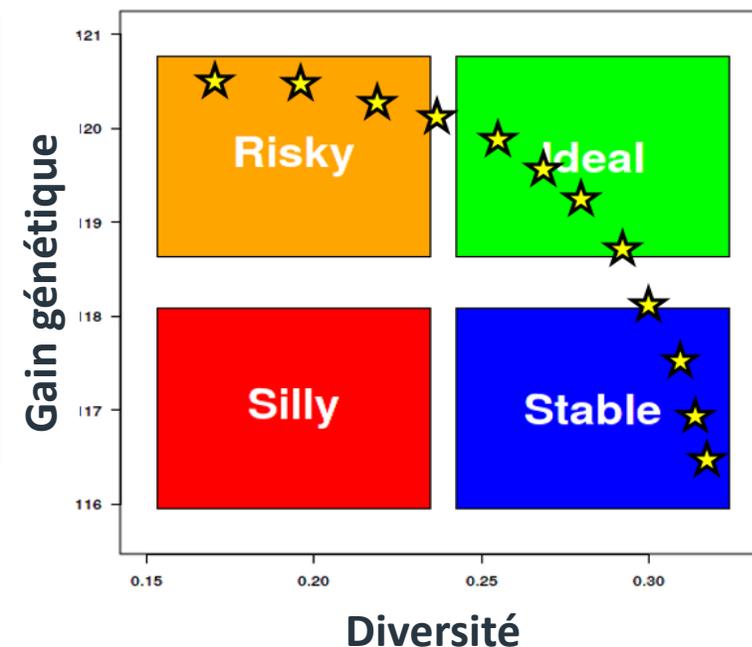
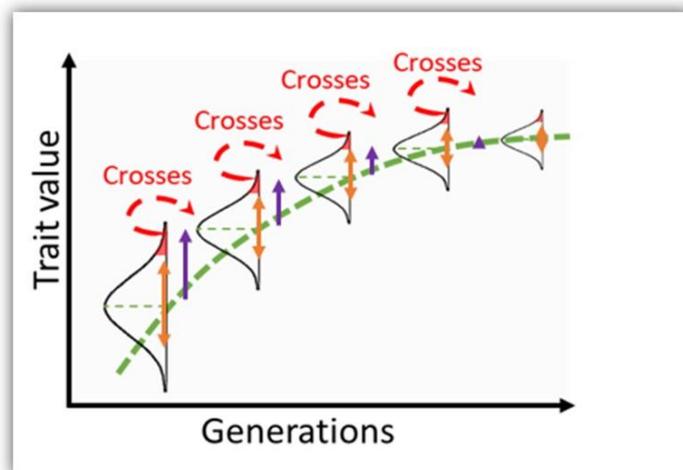
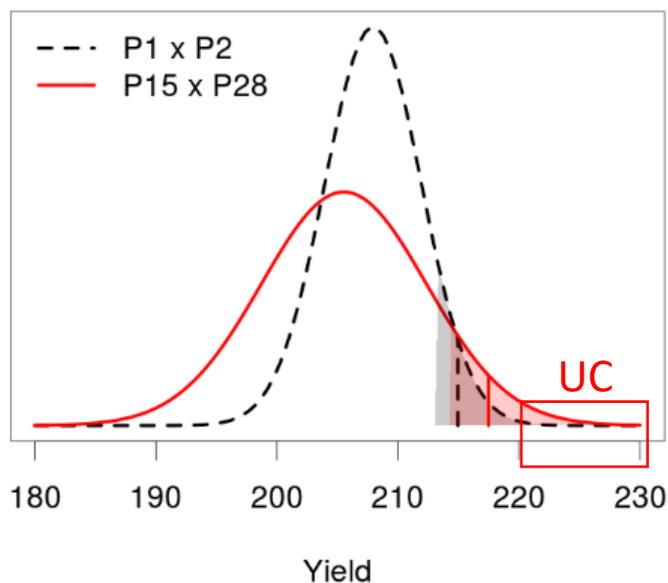
$$R = \frac{i \cdot \sigma \cdot r}{L}$$

$\sigma$ : Genetic Variance	$\sqrt{\sigma_G^2}$	Diverse starting population
$r$ : Accuracy		More information on varieties
$i$ : Selection intensity		Test more varieties
$L$ : Cycle length		Decrease time until re-crossing

Un choix plus éclairé des **croisements de départ** en simulant simultanément le progrès génétique possible à court terme et l'évolution de la diversité exploitable à plus long terme

## Trouvez les meilleurs croisements et ajustez la taille des familles

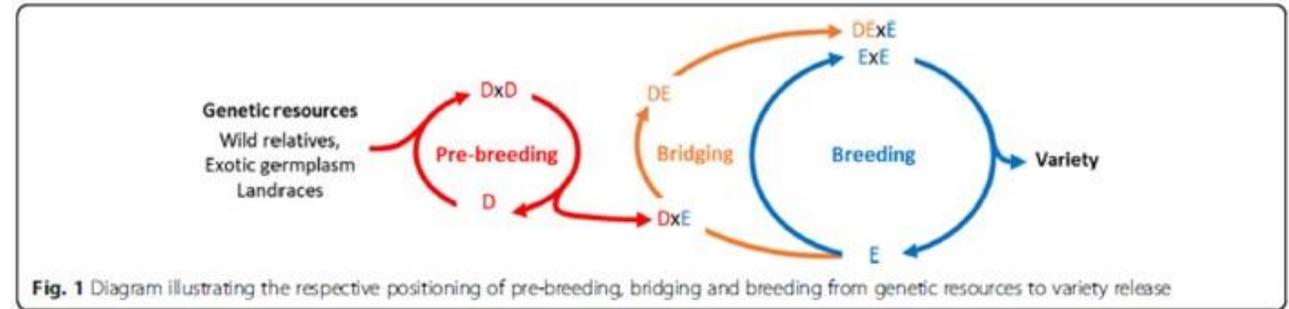
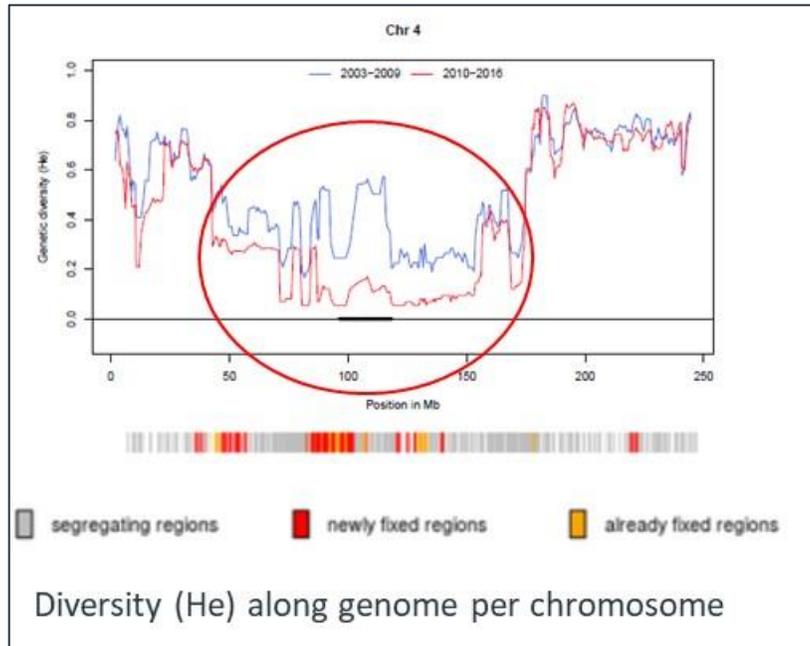
- La moyenne et la variance peuvent être prédites avec une plus grande précision
- Choisir des croisements produisant une bonne moyenne et une variance élevée (critères d'utilité)
- Optimiser le nombre de croisements et de descendance utiles



## Des applications sur la gestion et l’introduction de ressources génétiques

### Combiner Breeding et Pre-Breeding

- Identification des régions génomiques où l’introduction de la diversité est nécessaire
- Des croisements optimisés avec du matériel élite



université  
PARIS-SACLAY

Contributions to  
Genetic Diversity Management in  
Maize Breeding Programs  
using Genomic Selection

Thèse de doctorat de l’université Paris-Saclay

Ecole doctorale n° 581 : Agriculture, Alimentation, Biologie, Environnement, Santé (ABIES) Spécialité de doctorat : Sciences agronomiques  
Unité de recherche : Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France  
Réfèrent : AgroParisTech

Thèse présentée et soutenue à Gif-sur-Yvette, le 20 Janvier 2020, par

**Antoine ALLIER**

Composition du Jury

Christine DILLMANN Professeur, Université Paris-Sud	Présidente
Jean-Christophe GLASZMANN Directeur de recherche, CRAD	Rapporteur
Gregor GORJANC Associate professor, Roslin Institute	Rapporteur
Sophie BOUCHET Chargée de recherche, INRAE	Examinatrice
Leopoldo SANCHEZ-RODRIGUEZ Directeur de recherche, INRAE	Examinateur
Alain CHARCOSSET Directeur de recherche, INRAE	Directeur de thèse
Christina LEHERMEIER Responsable Statistical Genetics Unit, RAGT2n	Co-encadrante
Hélène PASCAL Sélectionneur maïs, RAGT2n	Invitée
Gwendal RESTOUX Chargé de recherche, INRAE	Invité

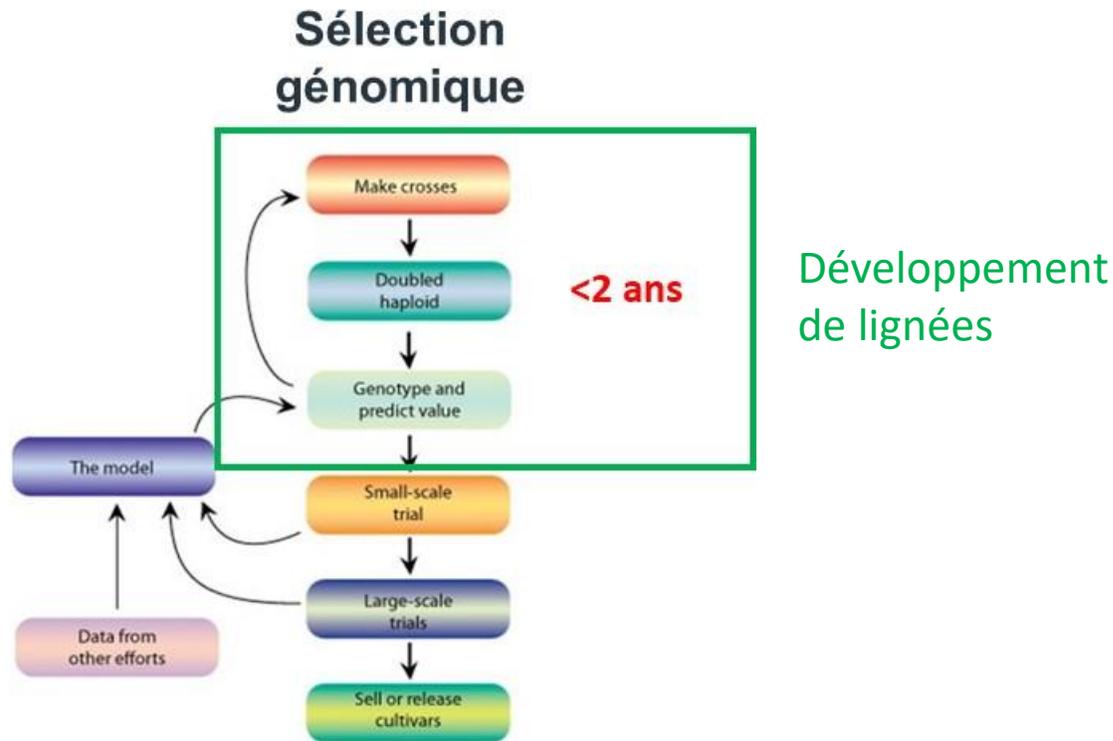
Thèse de doctorat  
NNT : 2020UPSA002

**Antoine Allier**, Simon Teyssède, Christina Lehermeier, Laurence Moreau & Alain Charcosset. Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genomics, 2020. DOI:10.1186/s12864-020-6756-0

## Sélection Génomique – encore plus d'intelligence dans le processus de décision (3)

Des protocoles de sélection rapide optimisés pour accélérer le progrès génétique : **la GS change la structure des schémas de sélection**

Une accélération du grain génétique par un **rebrassage rapide** pour concentrer les meilleurs allèles



### Prédiction ou évaluation de la descendance

- Décider quelles parties des populations doivent être phénotypés et quels individus peuvent être prédits
- Prédiction = aucune observation phénotypique sur certains caractères principaux (rendement)
- Choisir les individus qui peuvent être rapidement recyclés

### Le gain de temps est crucial

- Des plateformes de production de plantes HD
- Des plateformes SSD avec 3 générations / an
- Des équipements dédiés pour du speed breeding

$$R = i \cdot r \cdot \sigma_g / t$$

t : intervalle de temps  
entre générations

Time is  
genetic gain

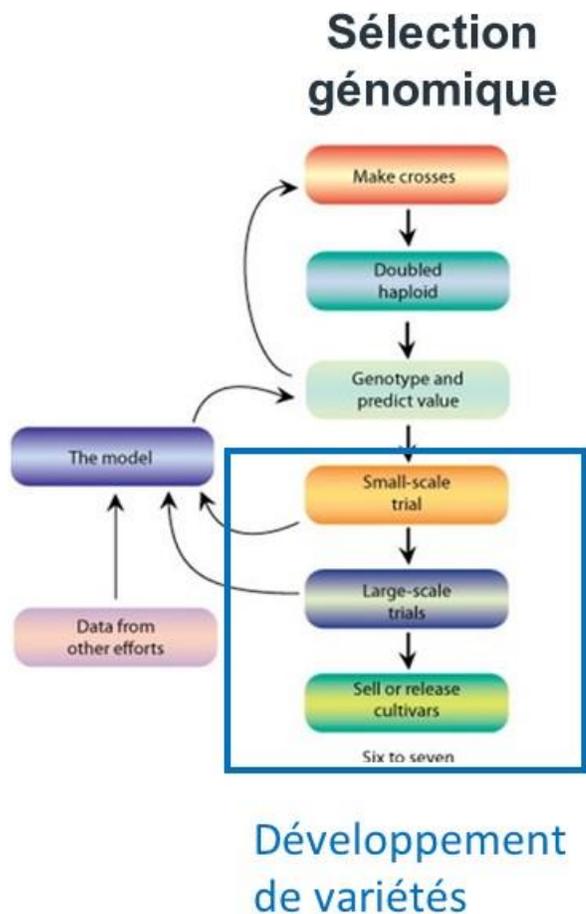
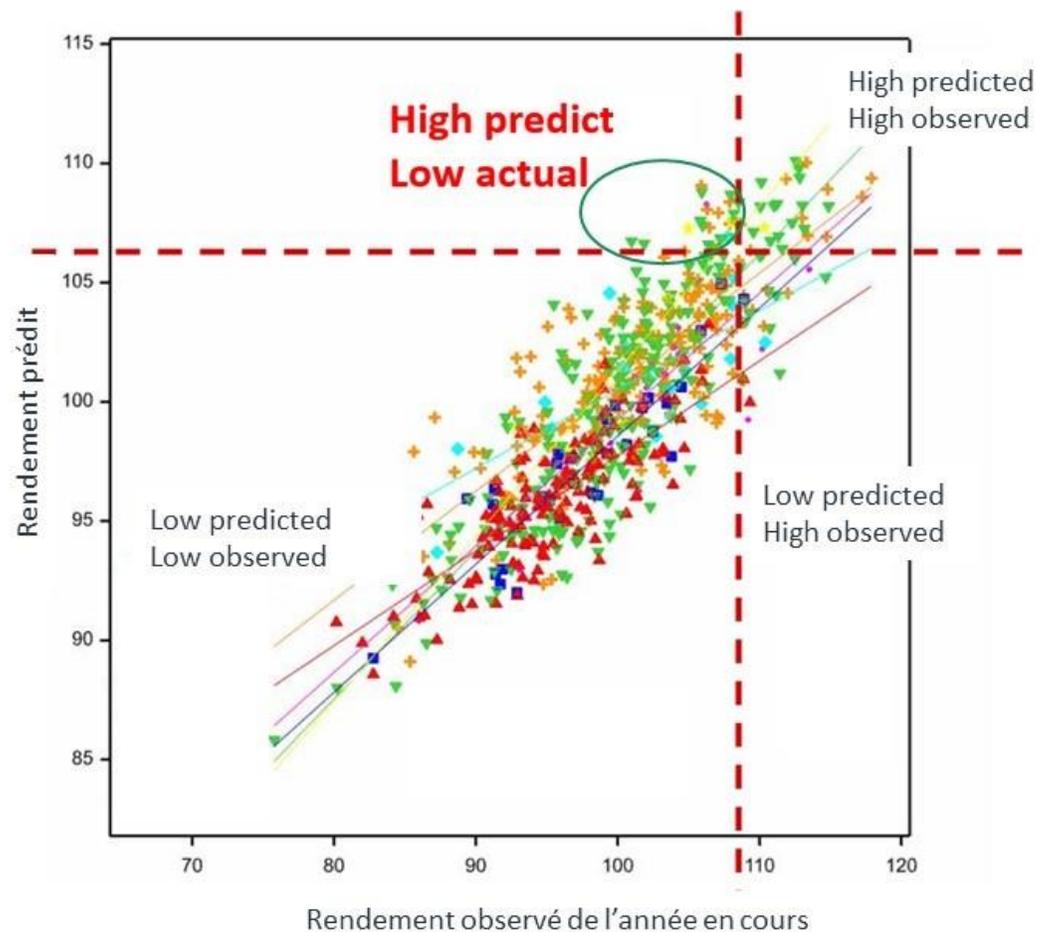
# Sélection Génomique – encore plus d'intelligence dans le processus de décision (4)

Un choix plus éclairé des individus à garder sur des modèles pluriannuels

Programme Blé France - Louville

Example of the use of GS in common wheat breeding programs

Predicted Performance vs. Observed Performance

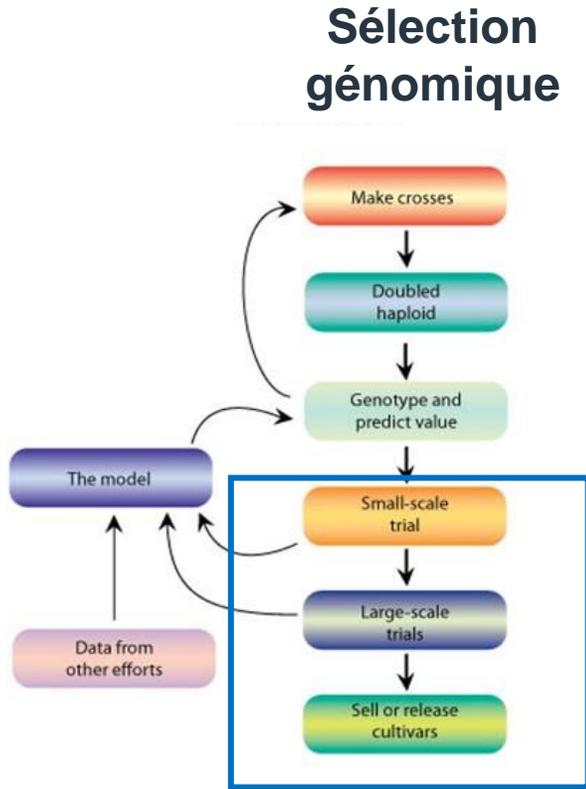


Une puissance accrue pour prédire des milliers de combinaisons Hybrides

## La GS pour prédire les hybrides non testés

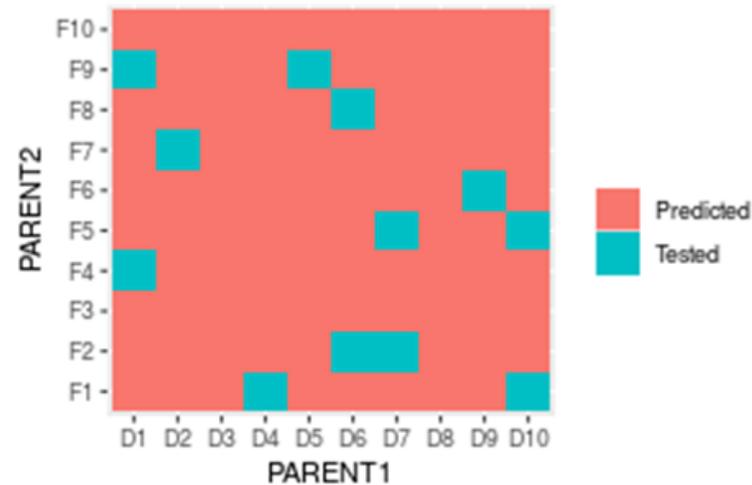
- Plans factoriels incomplets
- Exemple: Factoriel 100 x 100  
1 500 hybrides testés  
100 000 hybrides prédits

Optimise le choix des combinaisons à tester et à développer  
Aide à la création de pools hétérotiques



Développement de variétés

$$R = i \cdot r \cdot \sigma_g / t$$



université PARIS-SACLAY

Optimization of genomic selection for hybrids in a reciprocal selection program. Experimental evaluation and simulations on maize.

Optimisation de la mise en œuvre de la sélection génomique hybride dans un programme de sélection réciproque. Evaluation expérimentale et simulations chez le maïs.

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°581, agriculture, alimentation, biologie, environnement, santé (ABIES) Spécialité de doctorat: Sciences végétales Graduate School : Biosphera. Référent : AgroParisTech.

Thèse préparée dans l'UMR Génétique Quantitative et Evolution- Le Moullon (Université Paris-Saclay, INRAE, CNRS, AgroParisTech), sous la direction de Laurence MOREAU, Directrice de Recherche, le co-encadrement de Alain CHARCOSSET, Directeur de Recherche et la co-supervision de Gaëtan TOUZY, Cadre scientifique

Thèse soutenue à Paris-Saclay, le 08 novembre 2023, par

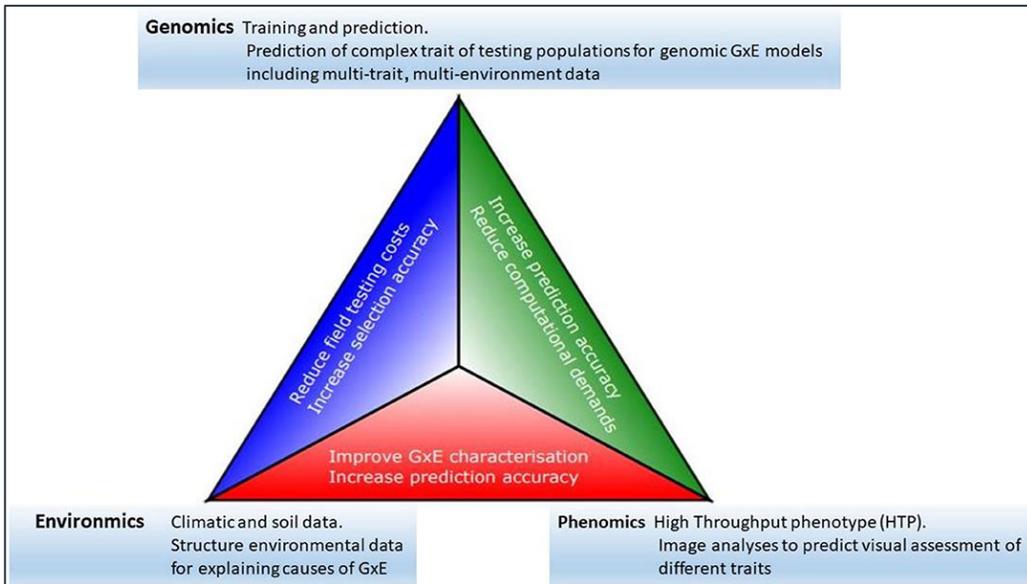
**Alizarine LORENZI**

Composition du Jury  
Membres du jury avec voix délibérative

Christine DILLMANN Professeure, Université Paris-Saclay	Présidente
Aaron LORENZ Professeur, Université du Minnesota (USA)	Rapporteur & Examineur
Leopoldo SANCHEZ Directeur de recherche, INRAE (centre Val de Loire)	Rapporteur & Examineur
David CROS Chercheur, Cirad (Montpellier)	Examineur

THESE DE DOCTORAT  
NNT: 2023UPAS0053

# Extension à d'autres technologies de phénotypage à haut débit



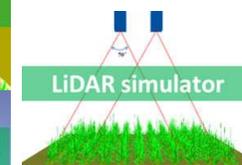
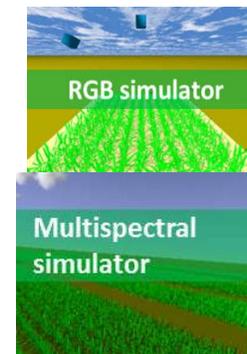
**Crossa J et al.** The Modern Plant Breeding Triangle :  
Optimizing the Use of Genomics, Phenomics, and  
Enviromics Data. Front Plant Sci. 2021



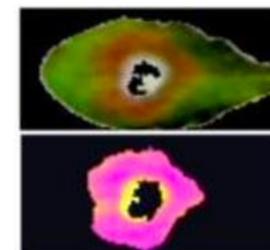
La méthode NAVAUTRON  
développée au LIPME pour  
mesurer des attaques de  
Sclerotinia Sclerotiorum sur le Soja

Augmenter la vitesse acquisition de données  
plus précises sans impacter les coûts

- Dans les essais terrains



- Automatisation des analyses de biotests



(Barbacci et al. 2020)



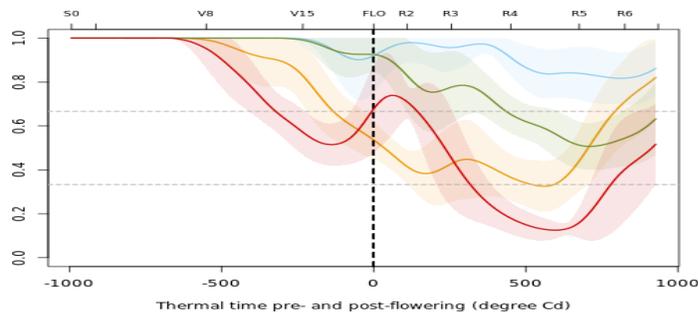
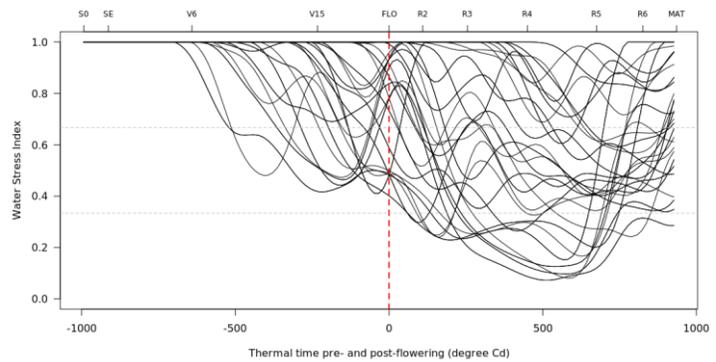
# Apprécier les effets environnementaux

Améliorer la précision du réseau d'essai (Envirotypage) en regroupant les lieux subissant des stress identiques (variation entre lieux et années).

$$\text{Phenotype} = \text{Genotype} + \text{Environment} + \text{GxE interactions} + \text{Error}$$

## Identifier des clusters d'Environnements (TPE)

- Optimisation du réseau d'essai (tester davantage dans les zones où l'héritabilité est plus faible)
- Evolution climatique : prédiction des risques climatiques pour la production/évaluation
- Identifier les caractères spécifiques à cibler (stress hydrique, tolérance au gel...)



Environment type ■ optimal ■ cold ■ heat and drought ■ drought and nitrogen

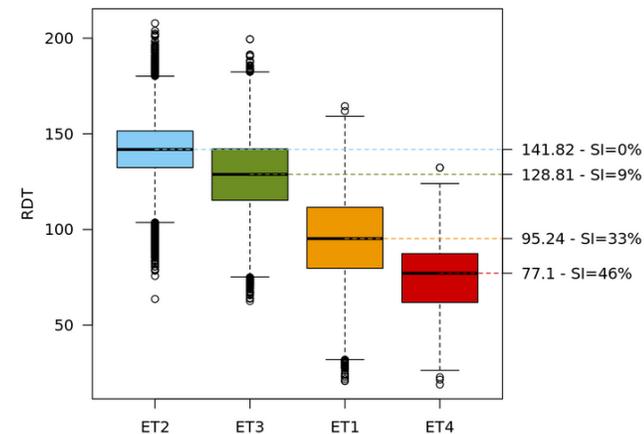
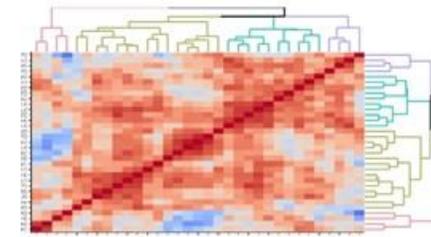
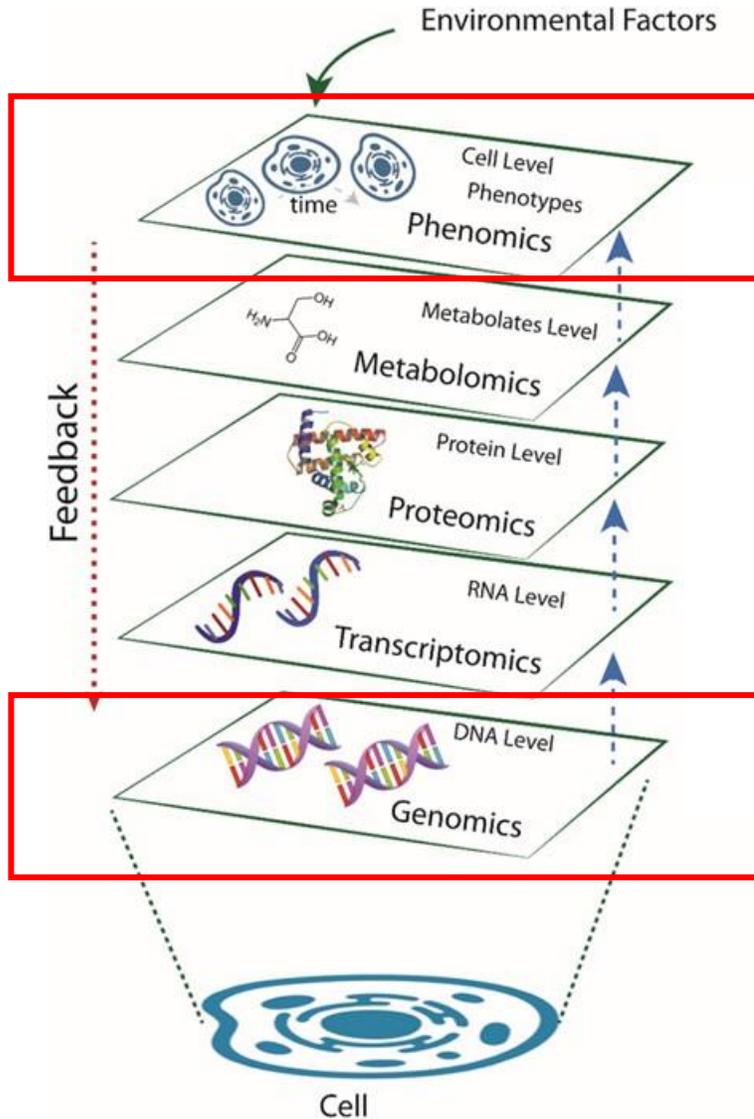


Figure 3.2: Yield performance according the drought cluster.



# AI et Prediction Multi-omics

- Explorer les données de séquençage des génomes ... pour identifier des gènes candidats et caractériser des réseaux de gènes



## Impact de l'AI sur la bioinformatique

**Shuang Peng & Loïc Rajjou.** Advancing plant biology through deep learning-powered natural language processing. *Plant Cell Reports*. 2024. doi.org/10.1007/s00299-024-03294-9

**Jingjing Zhai et al.** Cross-species modeling of plant genomes at single nucleotide resolution using a pre-trained DNA language model. *bioRxiv preprint* 2024. doi.org/10.1101/2024.06.04.596709

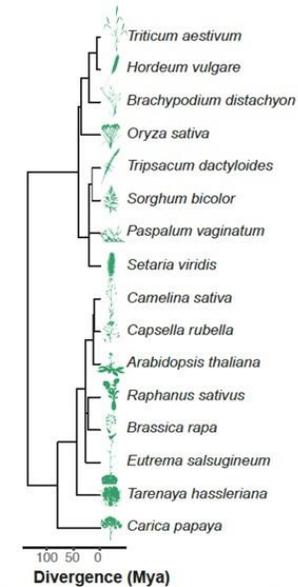
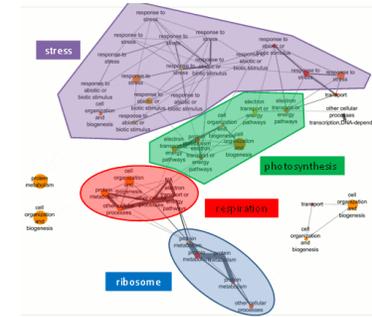
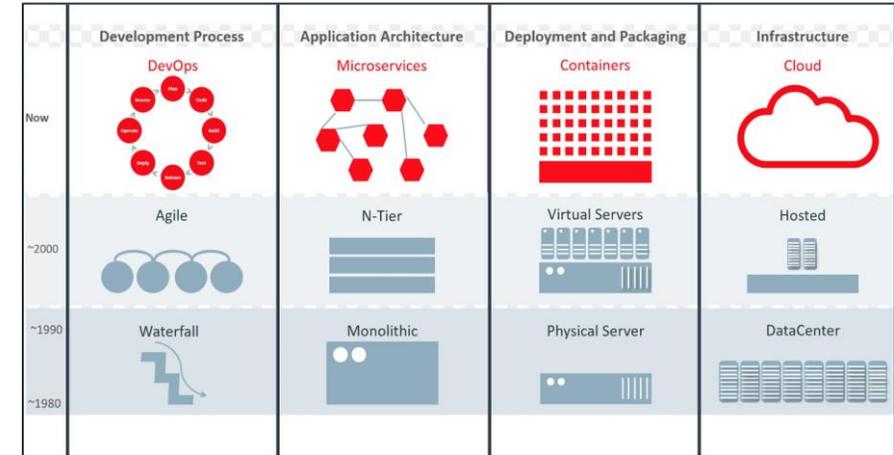
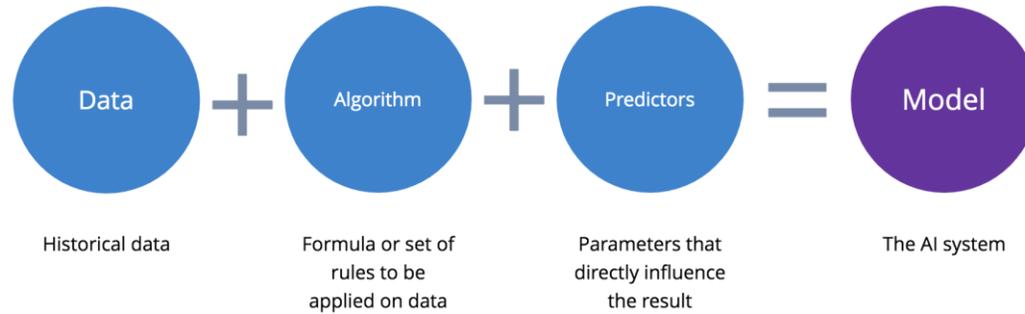


Fig 1. Overview of PlantCaduceus

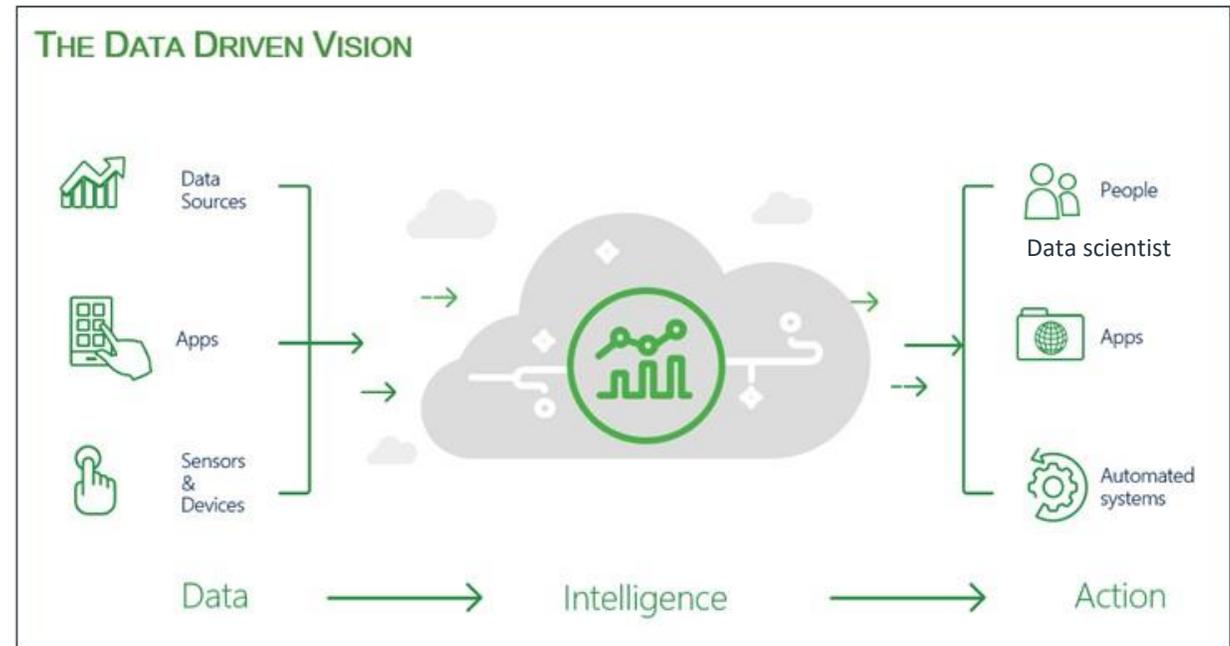


# Une mutation de l'entreprise vers une vision data driven



**La donnée au cœur des décisions**

what	period	data points
GENOTYPE	since 2000	milliards
PHENOTYPE	since 1985	millions
CLIMATE	since 2015	millions
GERMPLASM (lots)	since 2000	millions
GS - PREDICTIONS	in 2024	milliards



Le défi de la data et le Calcul Haute Performance (HPC)





# MERCI DE VOTRE ATTENTION !

## CONTACTS

[schatre@ragt.fr](mailto:schatre@ragt.fr) – Directeur Général RAGT2n

[pdufour@ragt.fr](mailto:pdufour@ragt.fr) – Coordinateur de recherche RAGT2n



think  
**SOLUTIONS**  
think **RAGT**