

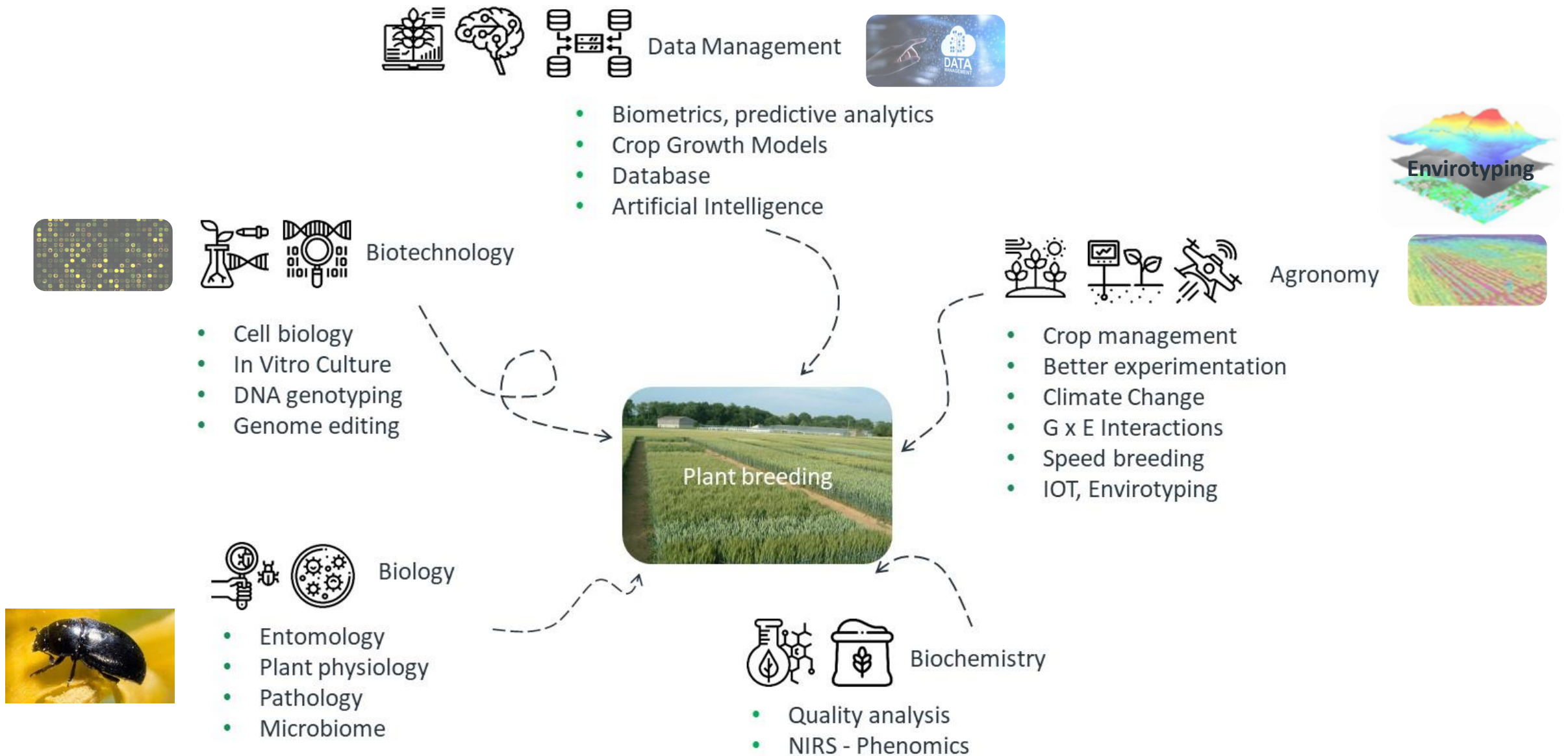
LA SELECTION VEGETALE DANS L'ERE NUMERIQUE

PHILIPPE DUFOUR
RAGT2N

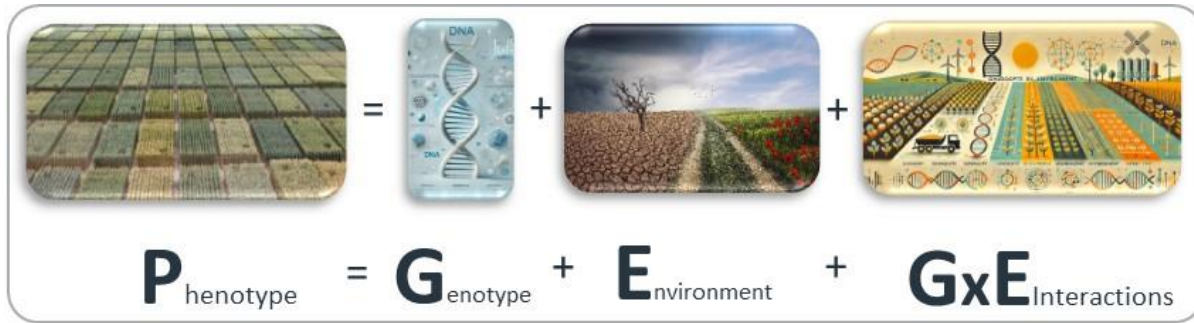


think
SOLUTIONS
think **RAGT**

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

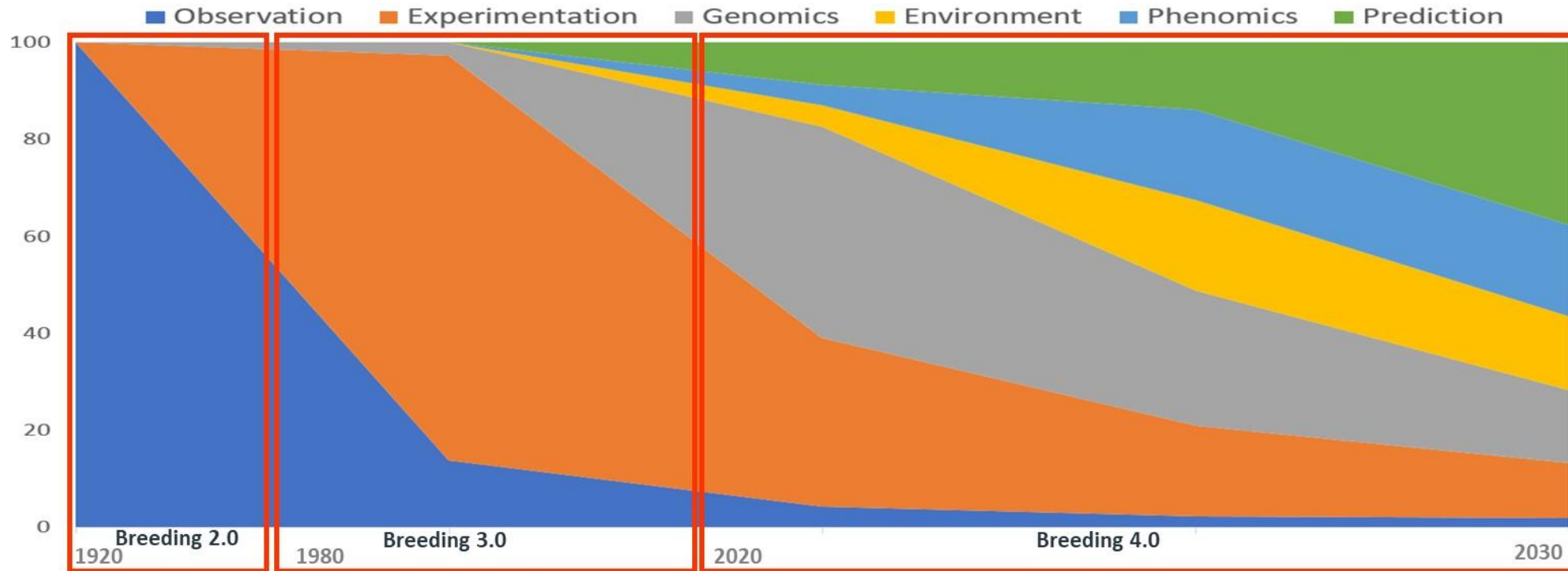


De la sélection phénotypique à la sélection prédictive

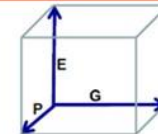
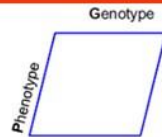


changements majeurs :

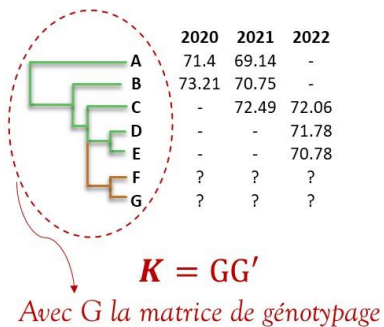
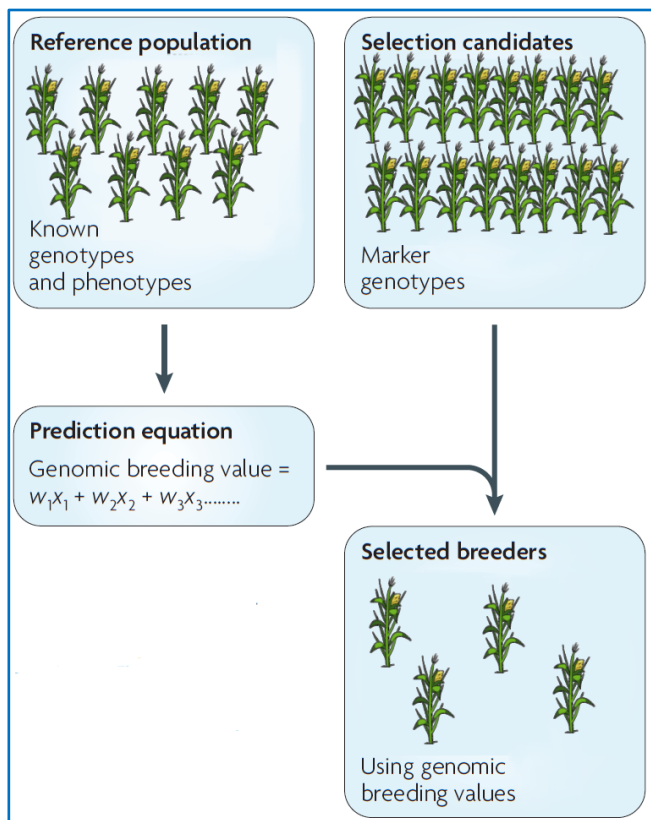
- Génotypage à haut débit
- Puissance de calcul / AI



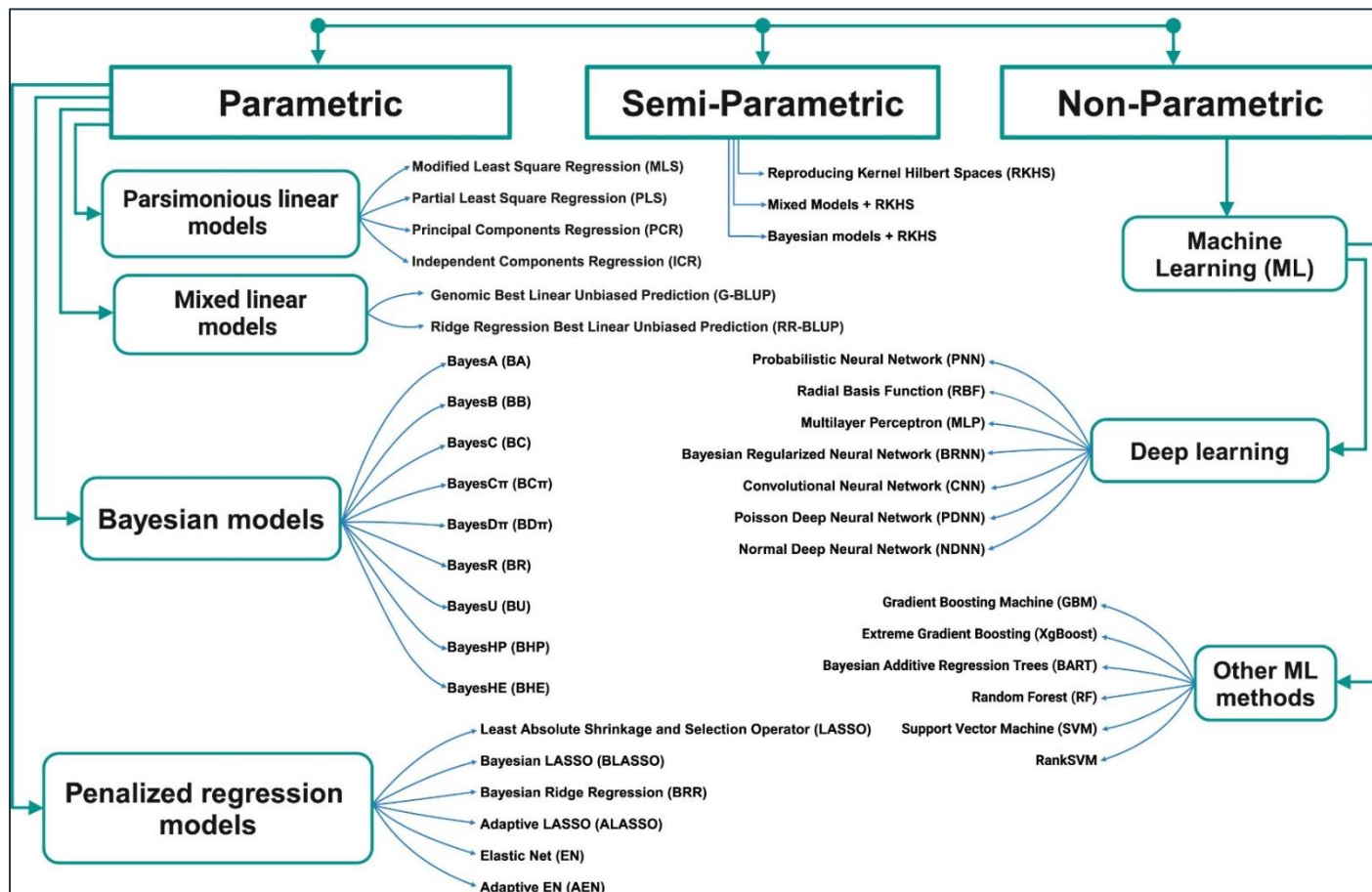
Phenotype



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



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

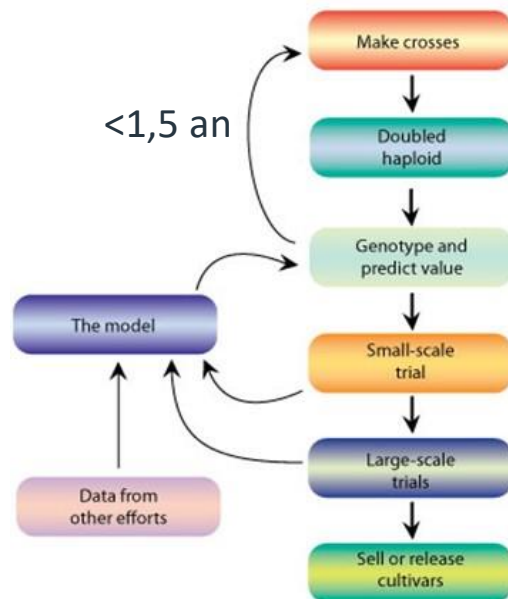
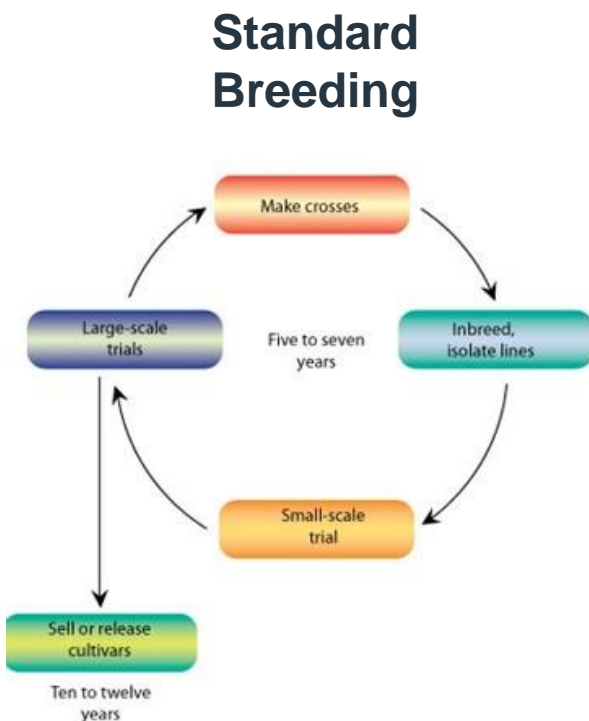


Le deep learning est particulièrement efficace dans les **tâches de prédiction**, où la quantité de données est énorme et où les relations entre les variables sont complexes et non linéaires, par exemple pour la reconnaissance d'image, l'estimation des effets GxE

Des nouvelles méthodes de sélection pour accélérer le gain génétique



Sélection génomique



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

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

σ : 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

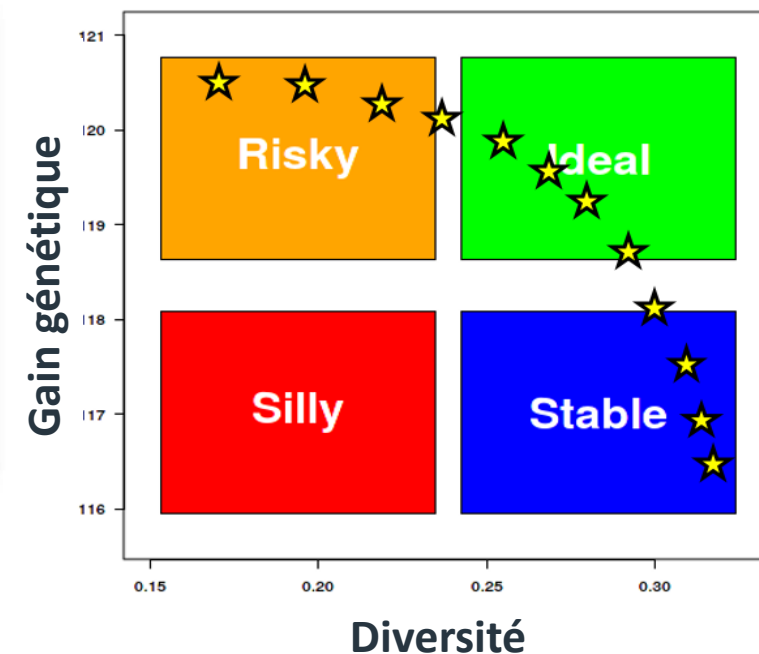
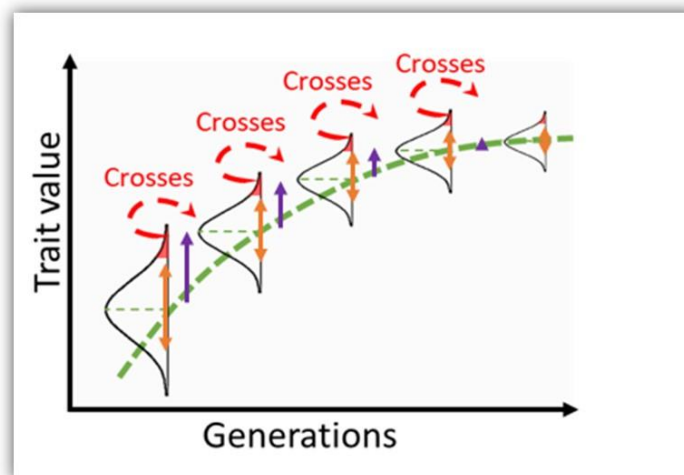
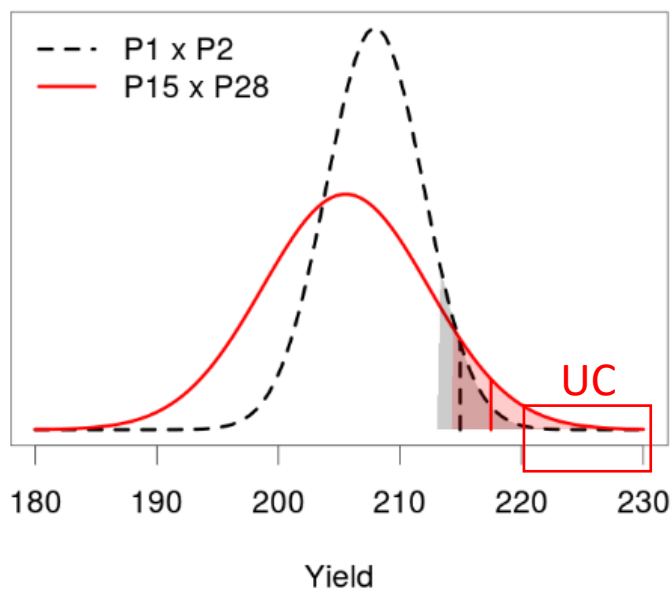
10 ans

< 6 ans

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

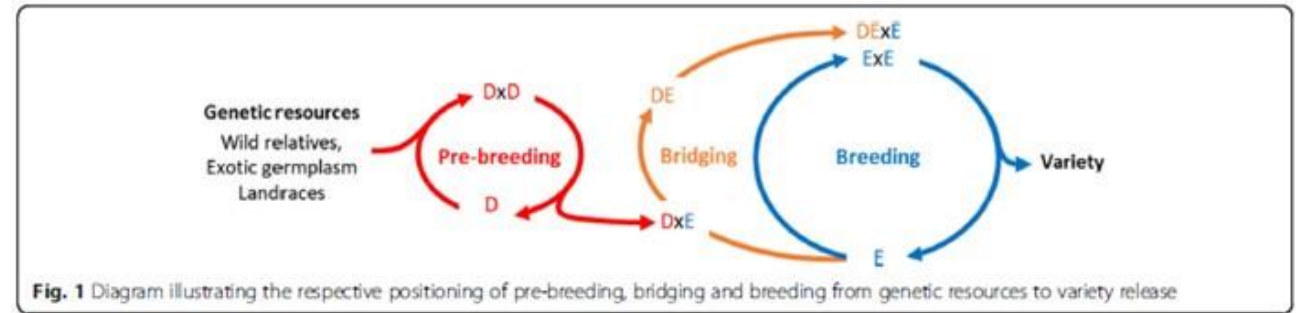
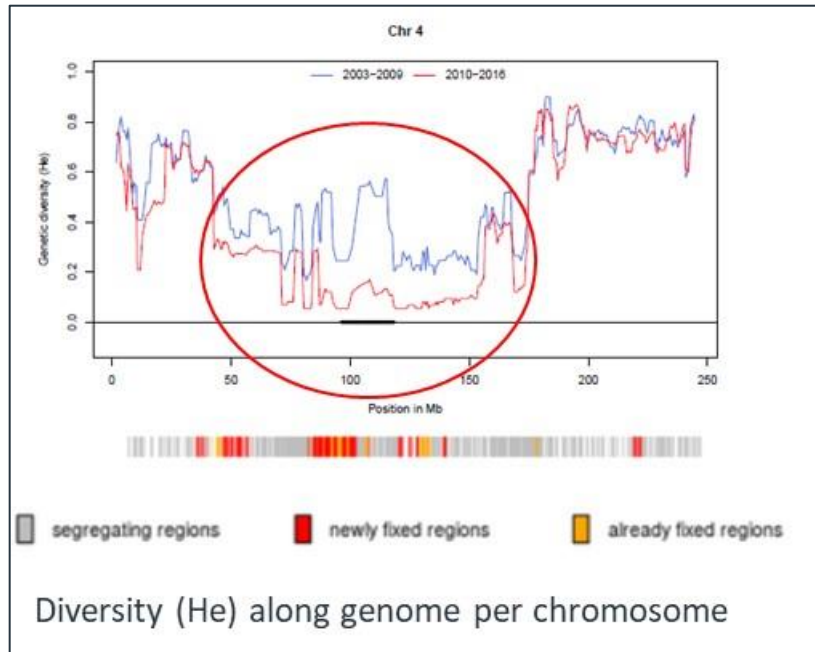
- La moyenne et la variance peuvent être prédites avec une plus grande précision
- Choisir des croisements produisant une bonne moyenne et une variance élevée (critères d'utilité)
- Optimiser le nombre de croisements et de descendance utiles



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, RAGTzn	Co-encadrante
Hélène PASCAL Sélectionneur maïs, RAGTzn	Invitée
Gwendal RESTOUX Chargé de recherche, INRAE	Invité

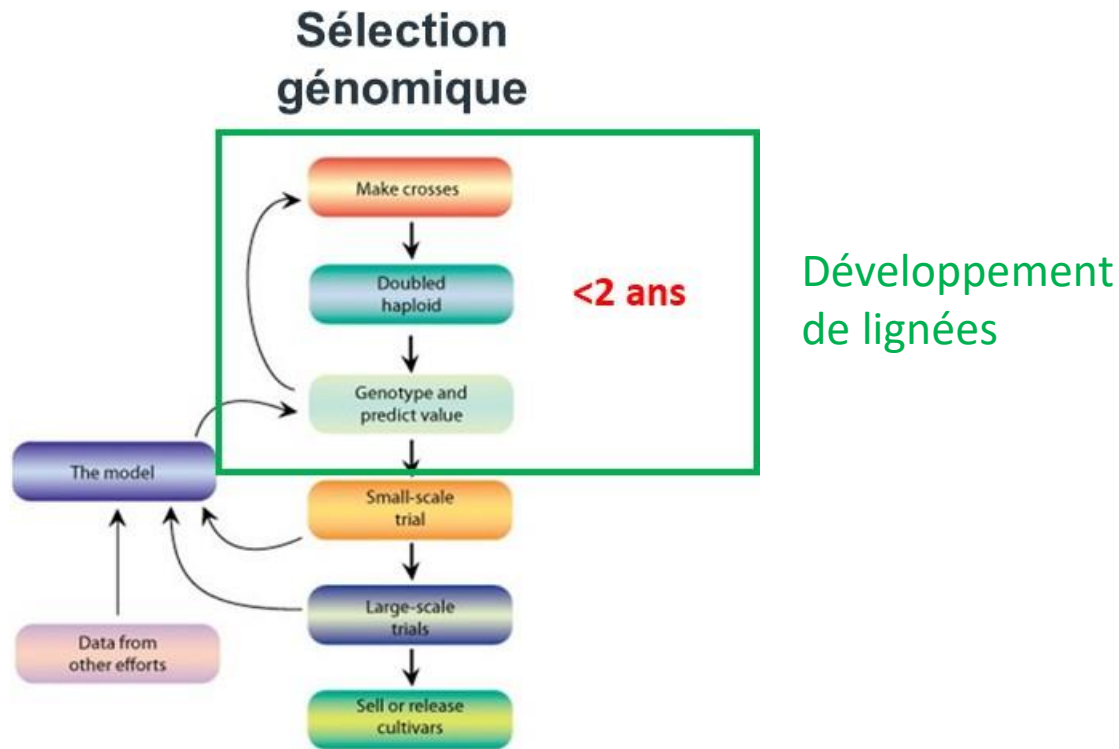
Thèse de doctorat
NNT : 2020UPSA002

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

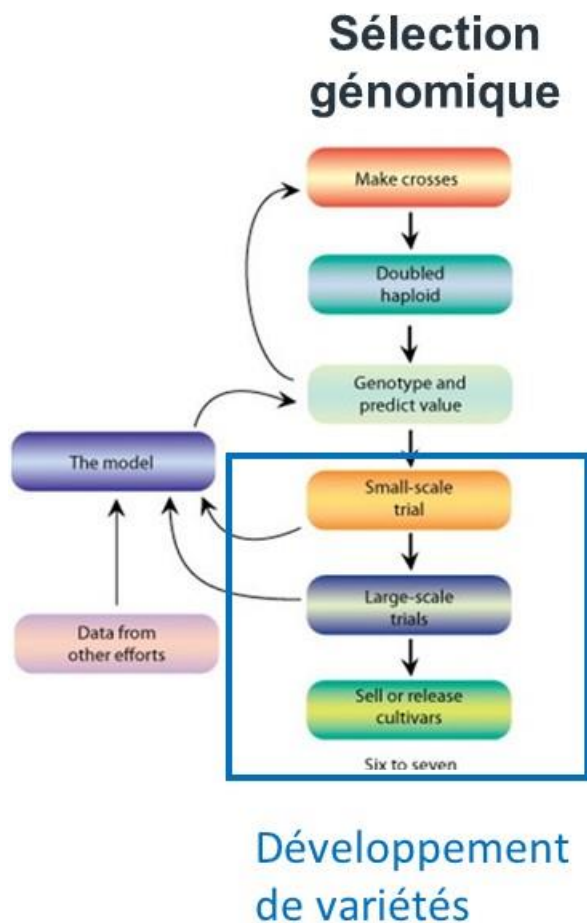
$$R = i \cdot r \cdot \sigma_g / t$$

t : intervalle de temps
entre générations

Time is
genetic gain

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

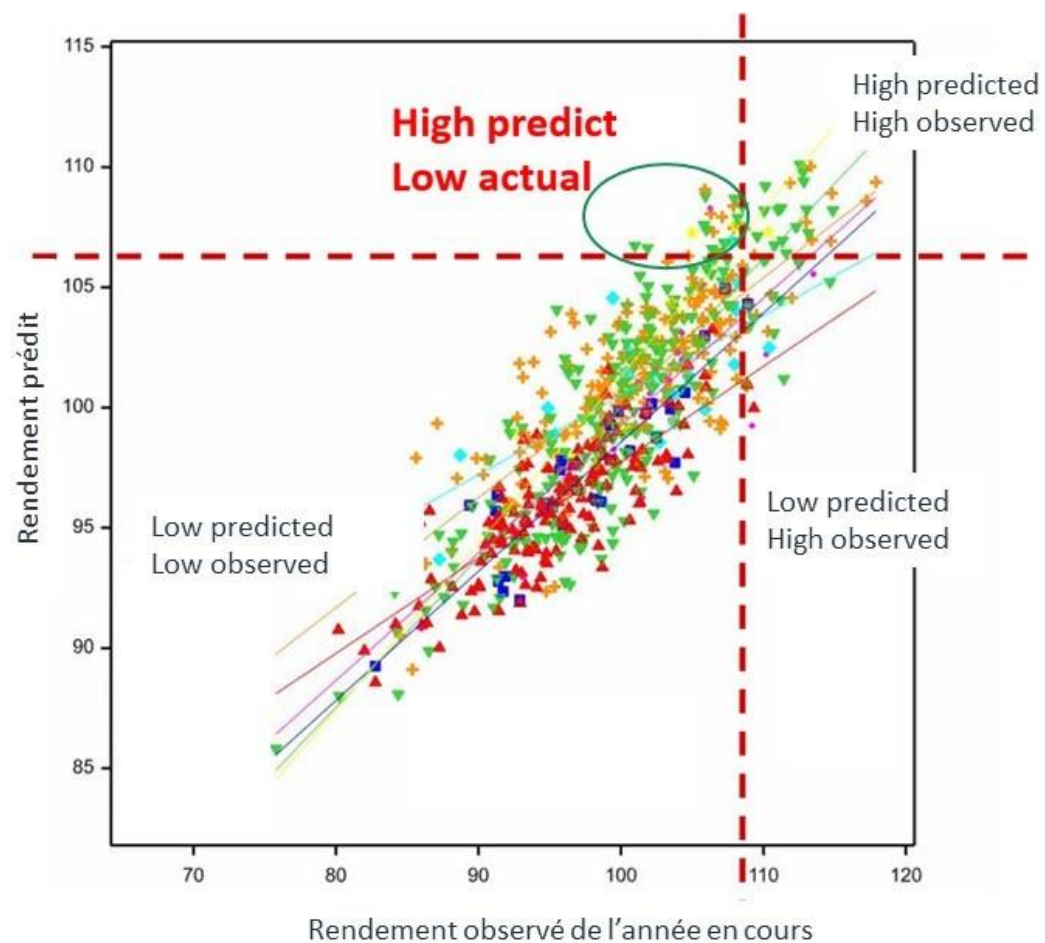
Un choix plus éclairé des individus à garder sur des modèles pluriannuels



Programme Blé France - Louville

Example of the use of GS in common wheat breeding programs

Predicted Performance vs. Observed Performance

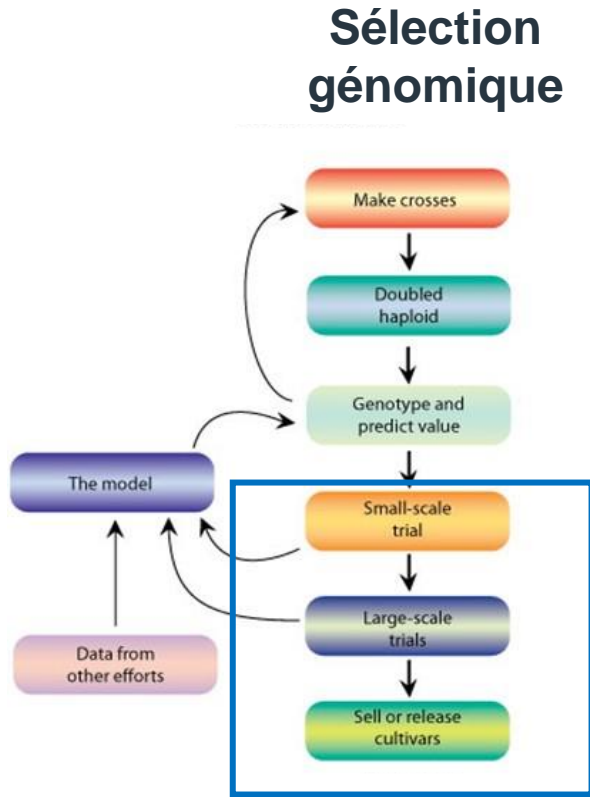


Une puissance accrue pour prédire des milliers de combinaisons Hybrides

La GS pour prédire les hybrides non testés

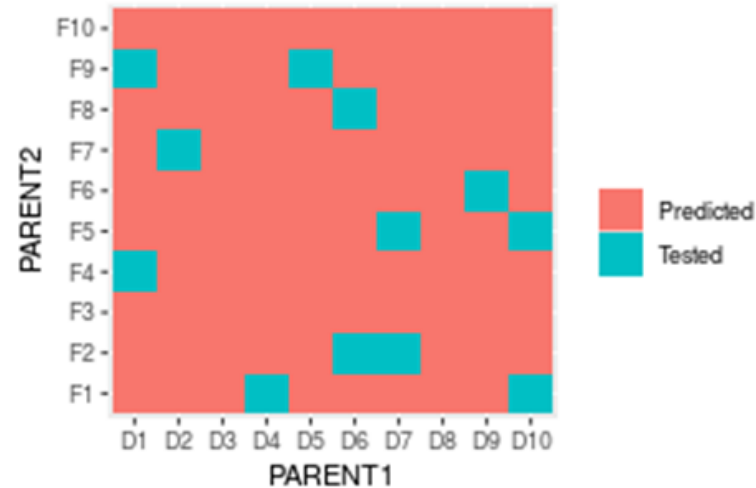
- Plans factoriels incomplets
- Exemple: Factoriel 100 x 100
1 500 hybrides testés
100 000 hybrides prédits

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

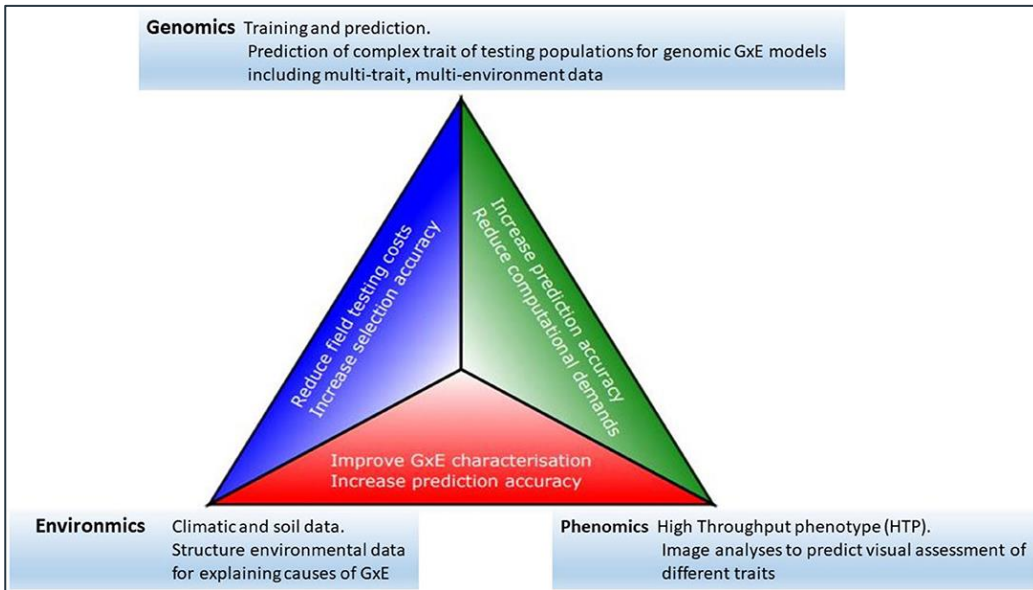
Alizarine LORENZI

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



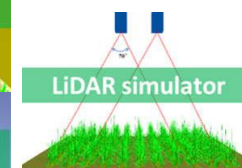
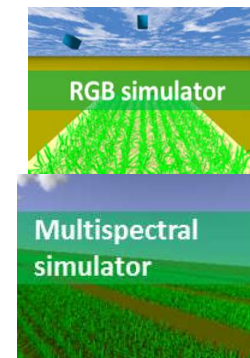
Crossa J et al. The Modern Plant Breeding Triangle :
Optimizing the Use of Genomics, Phenomics, and
Enviromics Data. Front Plant Sci. 2021



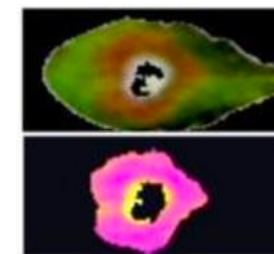
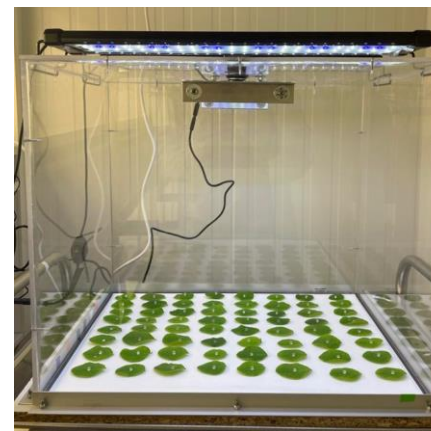
La méthode NAVAUTRON
développée au LIPME pour
mesurer des attaques de
Sclerotinia Sclerotiorum sur le Soja

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)



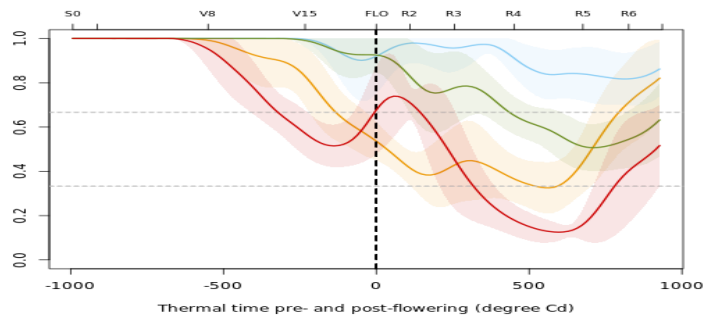
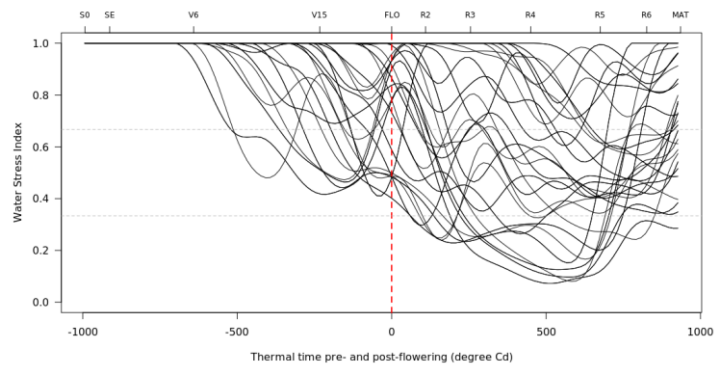
Apprécier les effets environnementaux

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

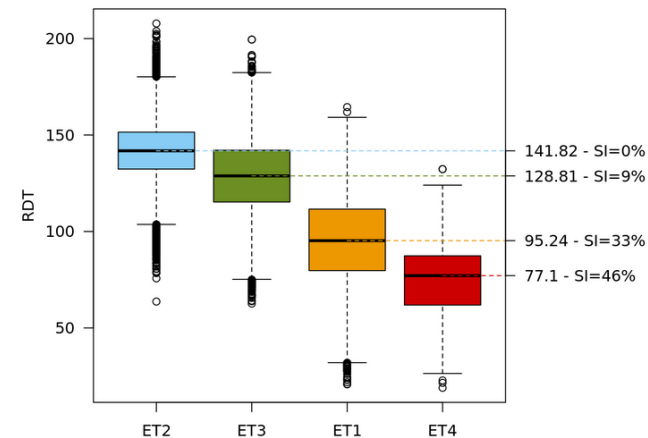
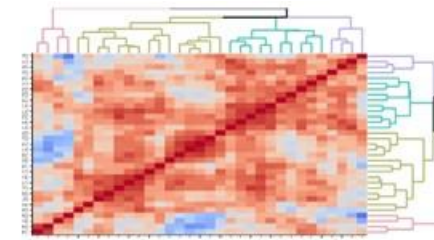
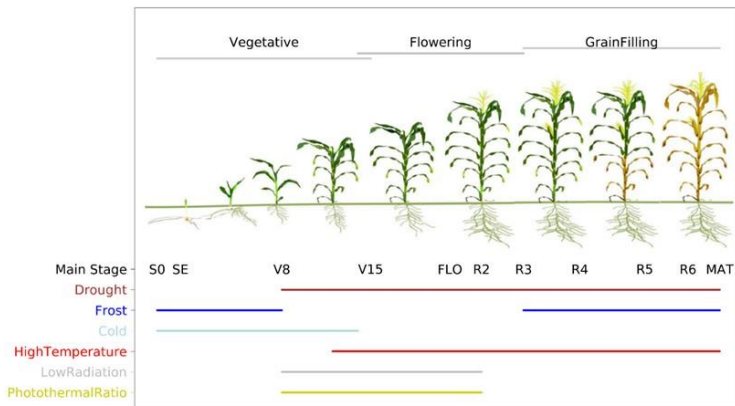


Figure 3.2: Yield performance according the drought cluster.

Apprécier les effets environnementaux pour une meilleure connaissance des interactions GxE

Caractères environnementaux pouvant être utilisés comme covariables pour une meilleure prédiction des performances

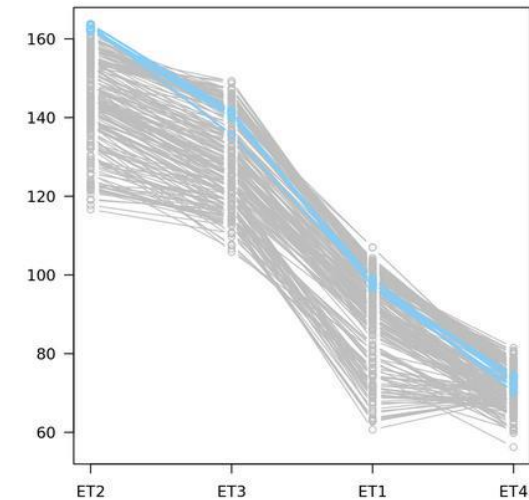
- Construire un modèle pour expliquer la variation du rendement en fonction d'indicateurs multi-stress
- Mieux prédire la stabilité des variétés dans chaque classe d'environnement



Utilisation de modèles de culture pour identifier des covariables environnementales dans chaque TPE

`nd_TEMP_MAX_sup35_FLO_R3`
nombre de jours où la température maximale est supérieure à 35°C de l'étage FLO à R3

Daniel Tolhurst et al. Genomic Selection Using Environmental Covariates Within an Integrated Factor Analytic Linear Mixed Model. 2021, DOI: 10.21203/rs.3.rs-995887/v1



AGC RDT

Environnements

Prédire des individus dans un environnement connu

Génotypes

0.21967075	0.49314384	0.34320554	0.41742623	0.63479006	0.76430785	0.11338207	0.78202641	NA	NA	NA	NA	NA
0.00142323	0.49310136	0.40629896	0.5000935	0.30731606	0.60938247	0.43828027	0.28648029	NA	NA	NA	NA	NA
0.48481728	NA	0.33696119	0.36669684	0.30775704	0.3202607	0.15220434	0.46139486	NA	NA	NA	NA	NA
NA	0.89862026	0.68812131	0.45064291	0.97676167	NA	0.51311177	0.49094605	NA	NA	NA	NA	NA
NA	0.8640784	0.57420481	0.29962097	0.1107422	NA	0.78136875	0.23064338	NA	NA	NA	NA	NA
0.54866334	0.75607257	0.03804197	0.47378265	0.36400204	NA	0.62451442	0.42203992	NA	NA	NA	NA	NA
0.57794267	NA	0.88162827	0.48797036	0.44801382	NA	0.23068139	0.63842683	NA	NA	NA	NA	NA
0.89139904	0.26681897	NA	0.45184097	0.14287995	NA	0.57997307	0.64532007	NA	NA	NA	NA	NA
0.87540484	0.19670798	NA	0.65401487	0.30131765	0.6561209	0.18429447	0.08412855	NA	NA	NA	NA	NA
0.41183998	0.33071623	0.96281513	NA	0.91108558	0.04902597	0.38525133	0.45438375	NA	NA	NA	NA	NA
0.40180113	0.59916991	0.4264862	0.09141099	0.65328363	NA	0.90026274	0.32735403	NA	NA	NA	NA	NA
NA	0.42112144	0.56508204	0.38784681	0.60442384	0.37272004	0.85700649	0.29665432	NA	NA	NA	NA	NA
NA	NA	0.09952485	0.12857067	0.10059293	0.98982204	0.99577501	0.69187531	NA	NA	NA	NA	NA
0.49045536	NA	0.99427387	0.14799	NA	0.4061467	0.9788799	0.92020066	NA	NA	NA	NA	NA
0.82232693	0.54721675	NA	0.9518	NA	0.78124512	0.37890509	0.32813344	NA	NA	NA	NA	NA
0.3817618	0.85221025	NA	0.9425	NA	NA	0.38674756	0.86033789	NA	NA	NA	NA	NA
0.7604827	0.0936478	NA	0.4604	NA	NA	0.67032875	0.61989507	NA	NA	NA	NA	NA
0.23404849	0.61010499	0.5772285	0.3465	NA	0.20740223	0.46526118	0.48100202	NA	NA	NA	NA	NA
0.66694973	0.4748825	0.40581169	0.15265316	0.05411087	0.16258311	0.40121464	0.98300051	NA	NA	NA	NA	NA
0.170898	0.3864386	0.71587276	0.54049953	0.90602919	0.981659	0.72961952	0.90260207	NA	NA	NA	NA	NA
0.601904	0.3778449	0.3334848	0.5117039	NA	0.84402912	0.71517136	0.07807603	NA	NA	NA	NA	NA
0.6891958	0.88930271	0.09029795	0.41919939	0.51441161	0.53363575	0.53254784	0.63682335	NA	NA	NA	NA	NA
0.0647082	0.88159796	NA	NA	0.11937284	0.77813757	0.33483795	0.03770021	NA	NA	NA	NA	NA
0.1427957	0.4005046	0.5220765	0.160031	0.18787627	0.59861712	0.08179556	0.32169182	NA	NA	NA	NA	NA
0.8483004	NA	NA	NA	NA	0.05295442	NA	NA	NA	NA	NA	NA	NA
0.9897246	0.7519197	0.1200637	0.6630854	0.27042021	0.36429485	0.74982666	0.82225046	NA	NA	NA	NA	NA
0.05866797	0.97761515	0.67440257	0.67094466	0.73463289	0.61237693	0.94921409	0.59936196	NA	NA	NA	NA	NA
0.0559909	NA	0.40020218	0.19711951	NA	0.23047892	NA	0.81245196	NA	NA	NA	NA	NA
0.57226983	0.99513929	0.41334133	0.37065386	0.77564457	0.5028024	0.98366224	0.79347506	NA	NA	NA	NA	NA
0.21479795	NA	0.91590265	0.02462978	0.69924673	0.45789403	0.53154599	0.25525202	NA	NA	NA	NA	NA

Prédire des individus dans un nouvel environnement

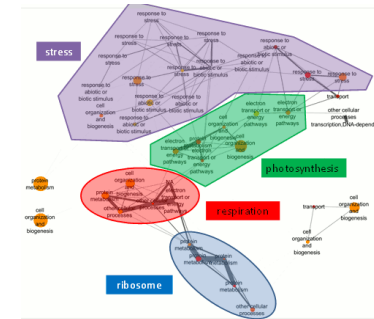
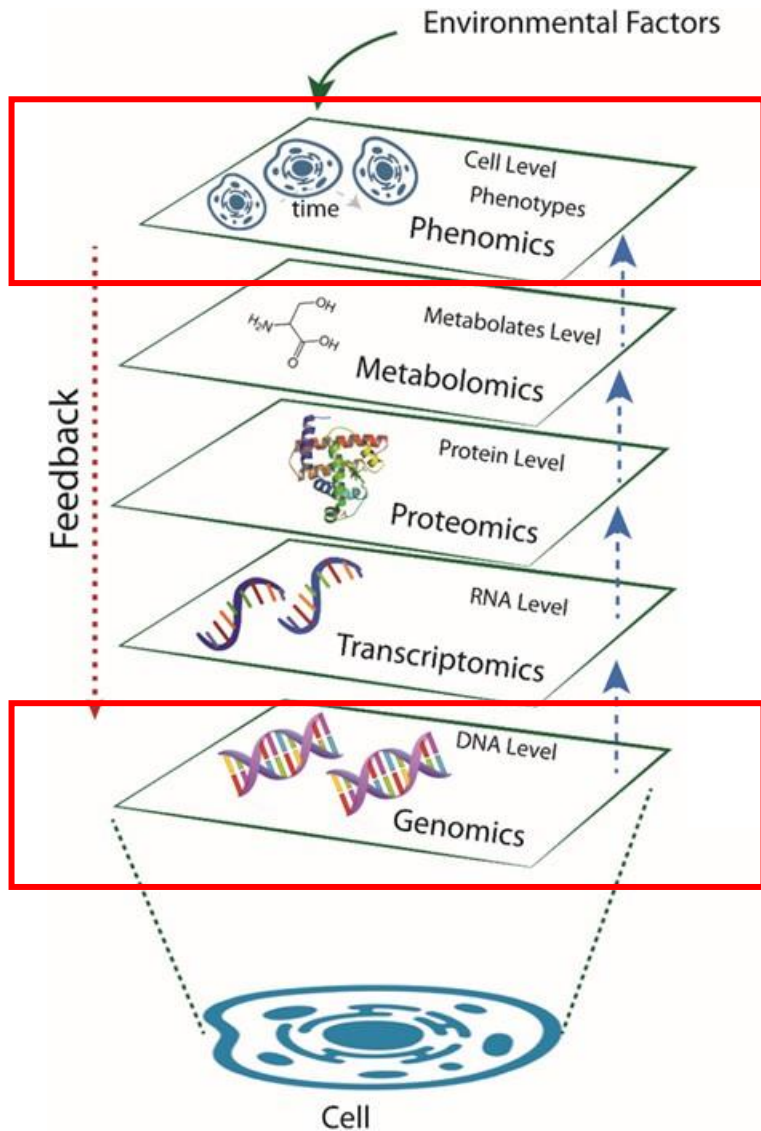
Prédire de nouveaux individus dans un environnement connu

Prédire de nouveaux individus dans un nouvel environnement



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

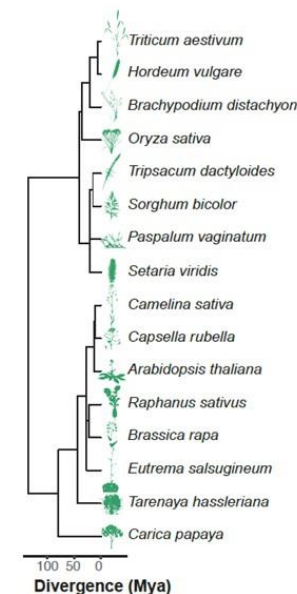


Fig 1. Overview of PlantCaduceus



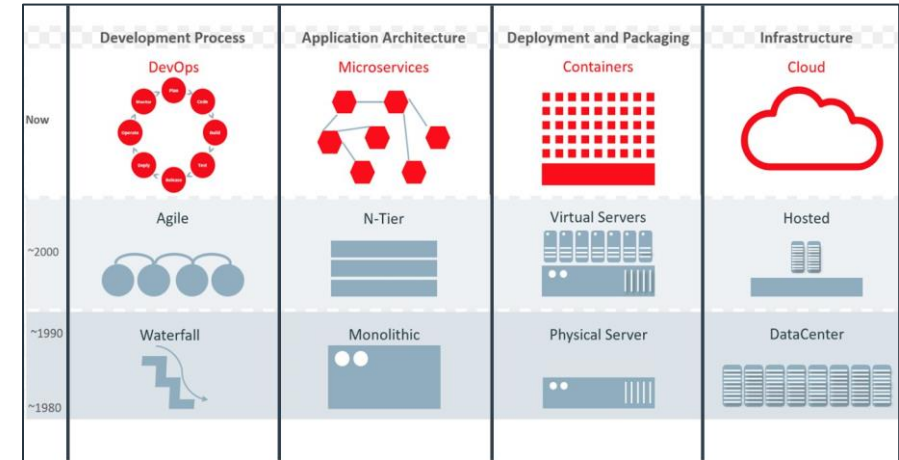
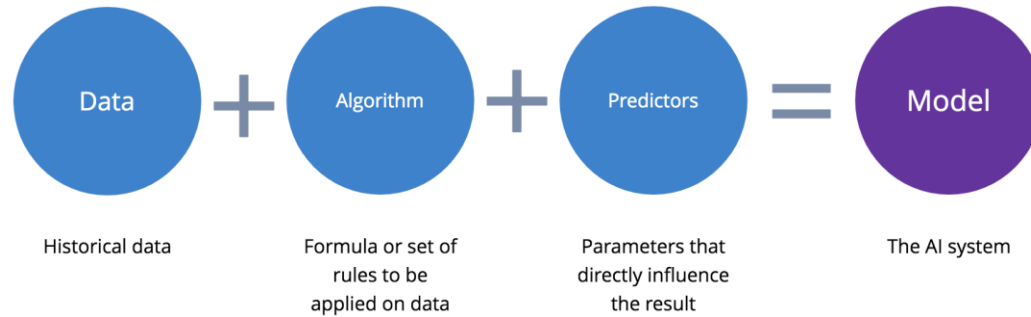
Lidea

FRAGT
2n

ARVALiS

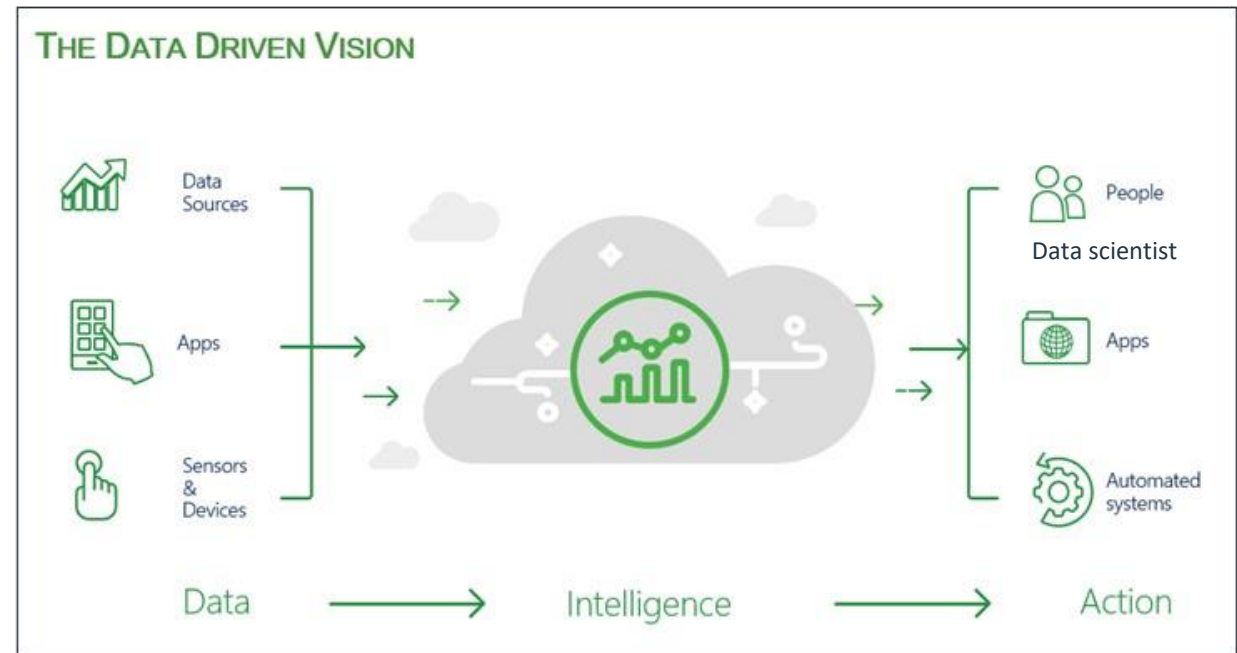
FRAGT
2n

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 – Directeur Général RAGT2n

pdufour@ragt.fr – Coordinateur de recherche RAGT2n



think
SOLUTIONS
think **RAGT**