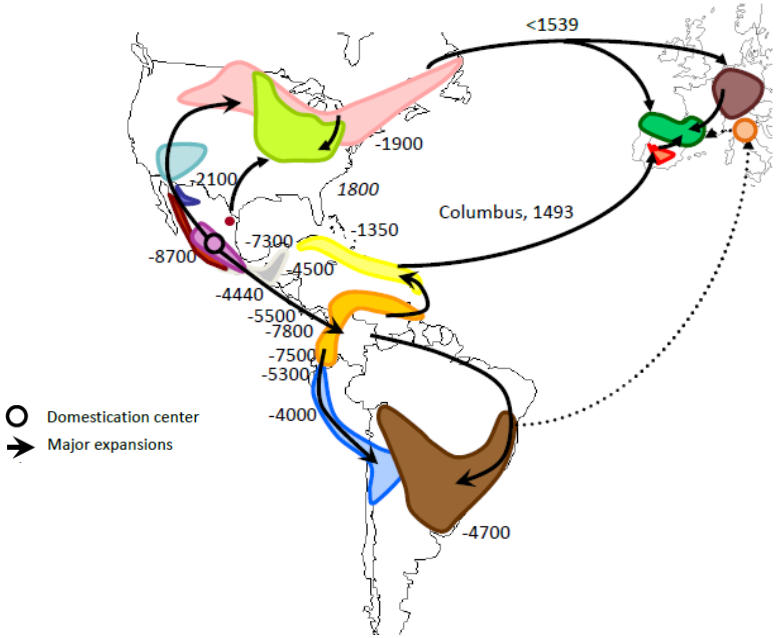
Use of genomic selection to better manage and enrich genetic diversity in breeding programs. Application to maize.

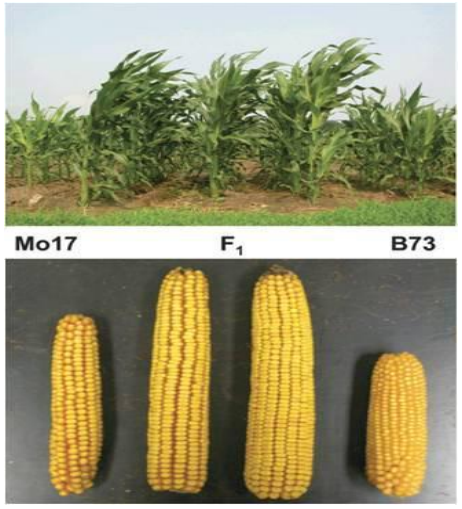
Laurence Moreau, Antoine Allier, Alizarine Lorenzi, Adama Seye, Stéphane Nicolas, Delphine Madur, Cyril Bauland, Christina Lehermeier, Simon Teyssèdre, Brigitte Gouesnard, Alain Charcosset

Maize breeding history

- Domestication in Mexico then adaptation to temperate climates
- Allogamous species, strong heterosis (Shull, 2014)
- 50ies : revolution landraces -> single-cross hybrids
- Formation of « heterotic » groups (some pre-existing some created by breeding) that now structure breeding material
- Competitive private sector, inbred lines from competitors not available for breeders (except « ex-pvp »)



Tenaillon and Charcosset, 2011



Springer and Stupar, 2007

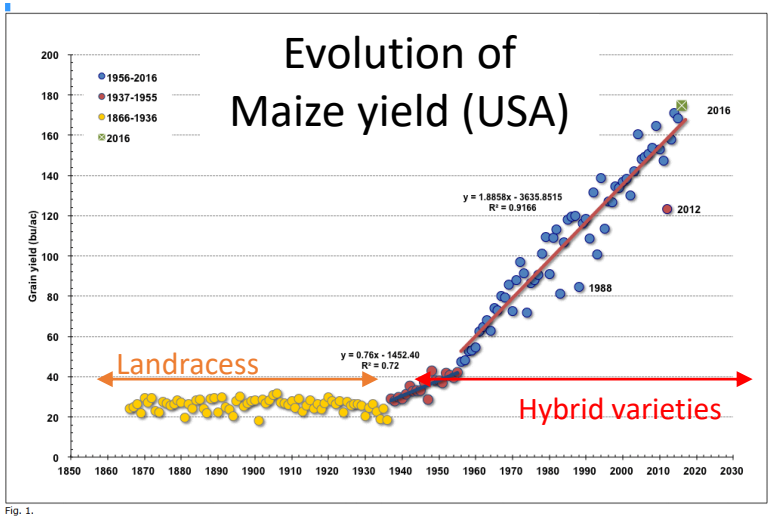
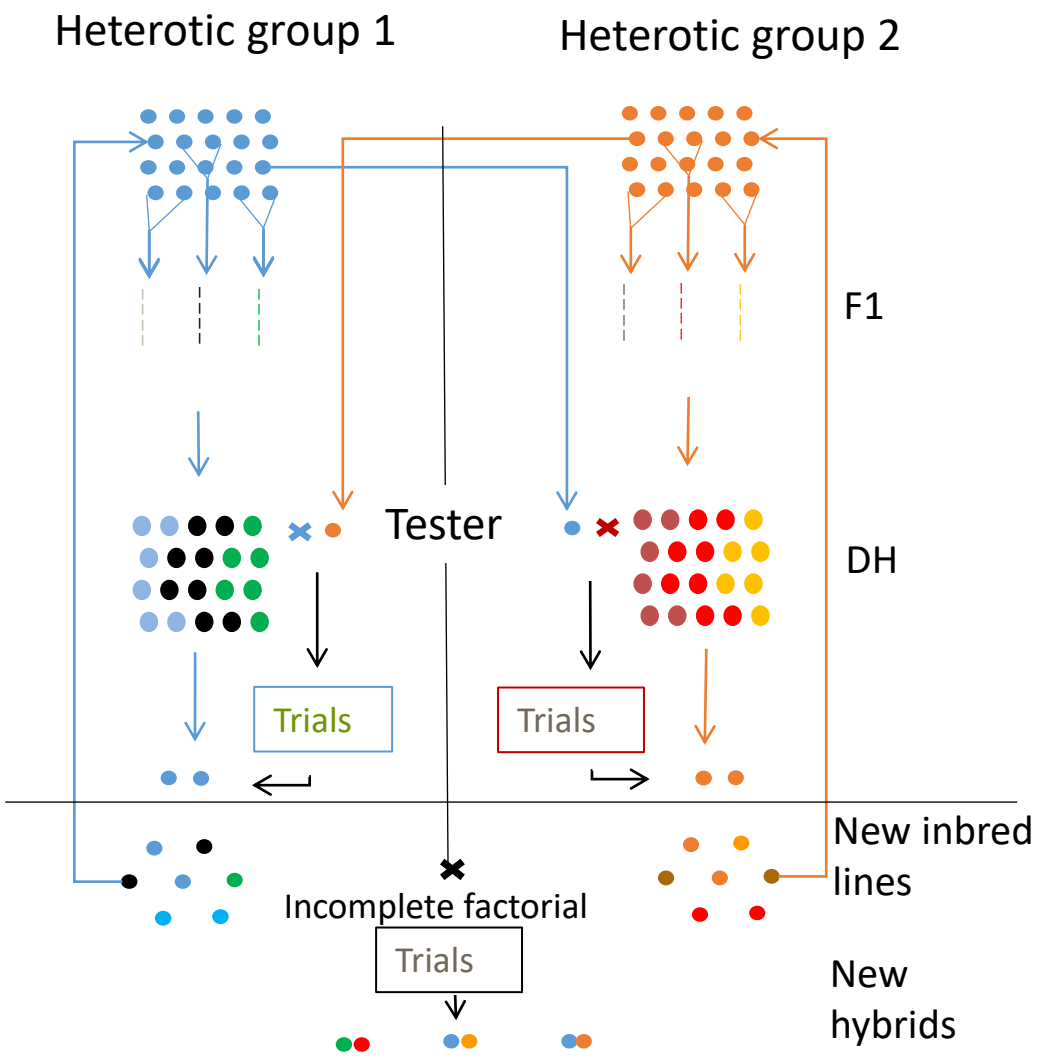


Fig. 1.

⇒ **Old population varieties (Landraces) : sources of genetic diversity only partially exploited when breeding moved to hybrids but which show large performance gap / elite**

Reciprocal recurrent selection



Selection in each heterotic group for the hybrid performance of candidate lines when crossed to lines from the other group

In each group, candidate lines now mostly produced by **haplo-diploidization** to accelerate their creation (2-3 generations / 8 generations with selfing)

To facilitate selection, evaluation of candidates DH made in two stages

1. Candidates are crossed to a limited number of lines (testers) from the other group **selection on the General Combining Ability (GCA)**

2 Only the best candidates are crossed in a partial factorial to identify the best hybrids **selection on GCA + Specific CA**

⇒ **« Long » process** (8 years with SSD, 5 years if DH / cycle), **expensive** (phenotyping of the two stages) that only **partially exploits the complementarity** between groups

« Genomic Selection » a breakthrough innovation in breeding

Principle of GS:

Thanks to the development of cheap and dense genotyping approaches use of all markers as predictors of the genetic value of (non –phenotyped) individual.

Two-steps:

(1) Calibration of the model in a training population (phenotypes + markers)

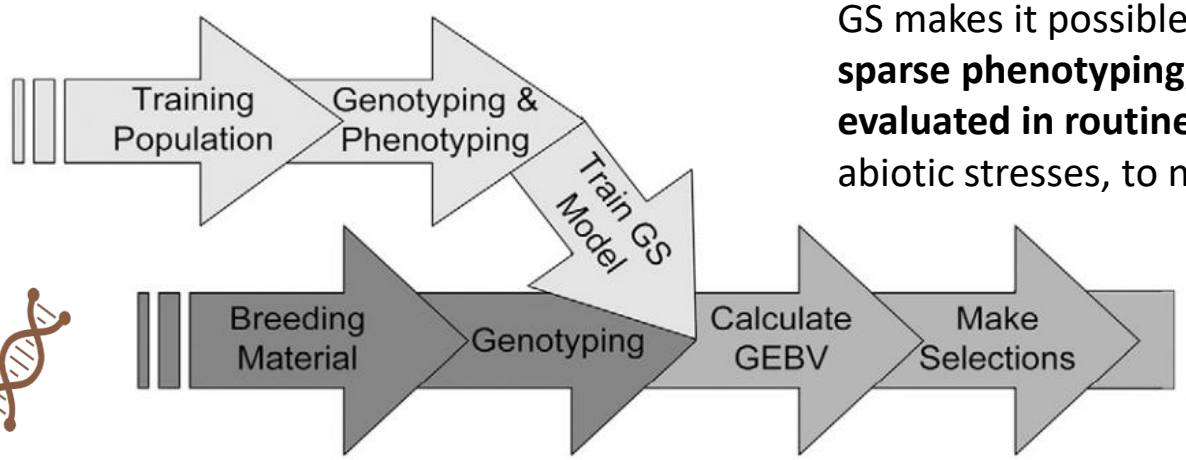
(2) Predictions in selected population based on markers only.

1 +

Phenotype (Y) Genotype (X)

2 Marker effects a_m +

Genotype (X)



GS makes it possible to perform **early selection**, **sparse phenotyping, select for traits that cannot be evaluated in routine** (adaptative traits to biotic or abiotic stresses, to new farming practices…)

$$G = \sum_m a_m x_m$$

Important to define an adapted training set !

(Optimization criteria available. ex Rincent et al. 2012, Rio et al. 2022)

Use of GS in maize breeding programs

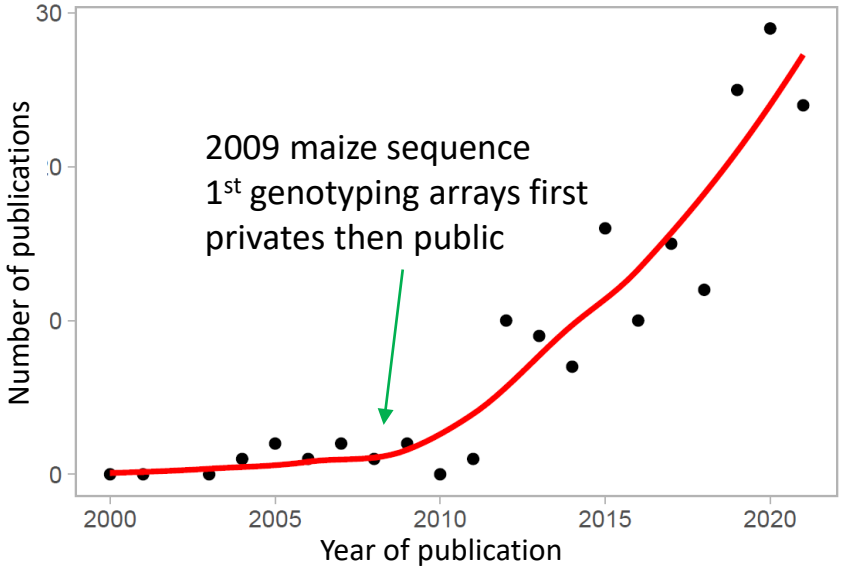
- GS quickly tested by maize breeders to reduce phenotyping in biparental progeny using low density genotyping techniques

Ex: Johnson, Plant Breeding Reviews, vol 24, **2003.**
Results obtained in a F2 progeny with 89 markers illustrating the interest of GS predictions over QTL-based predictions

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

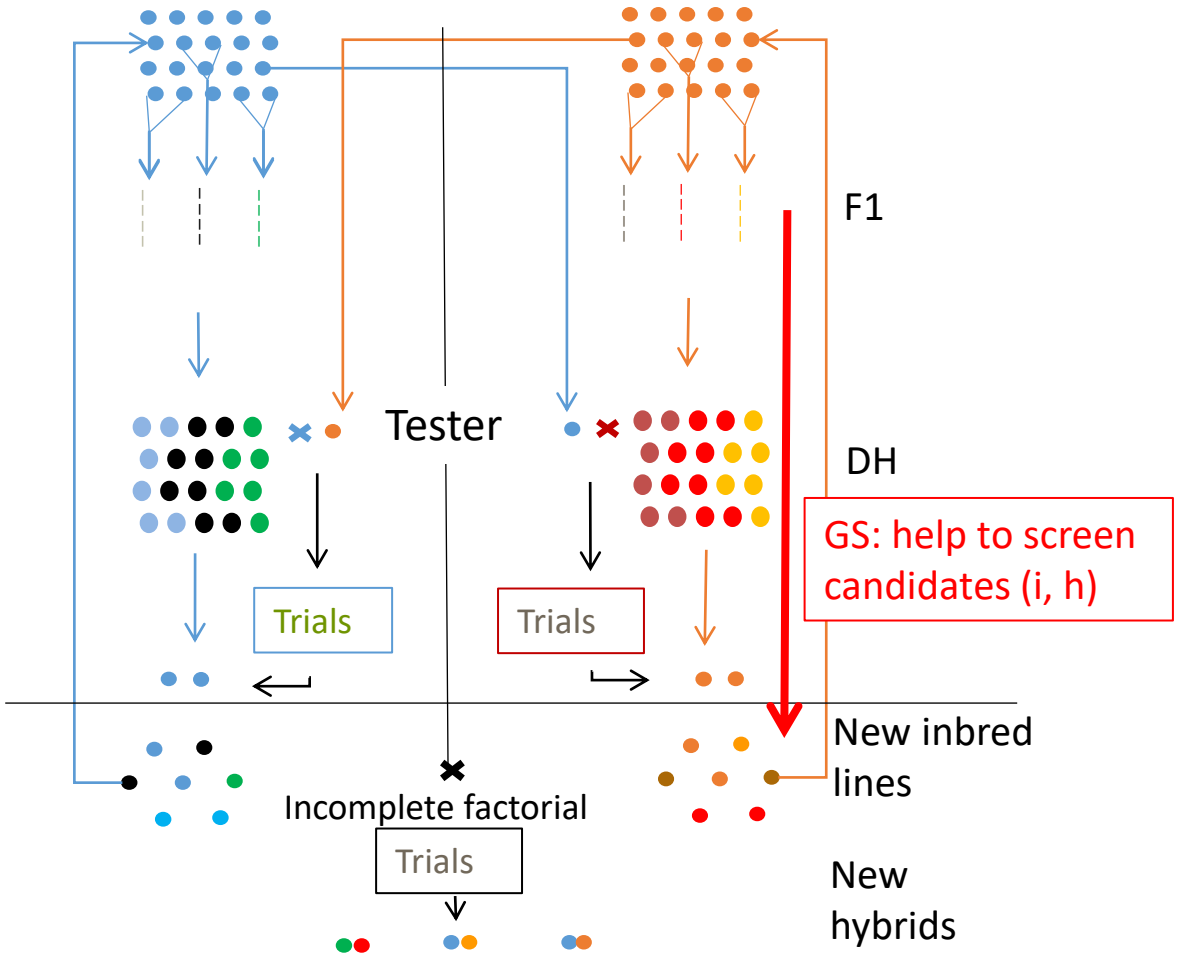
- 2009 maize sequencing and first genotyping arrays
- Today **genotyping costs** = 10-12 euros for thousands of markers = **cost of one field plot**
- GS now adopted by most (all ?) maize breeding companies
- Large number of publications per year on the use of GS in maize

Publications on Genomic selection in Maize

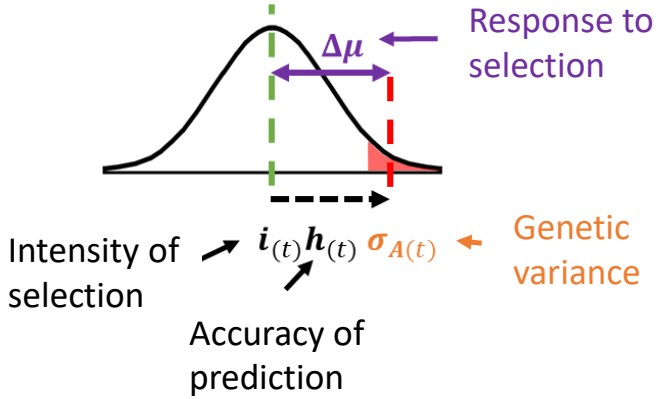


Source :Web of science

Use of GS in maize breeding programs

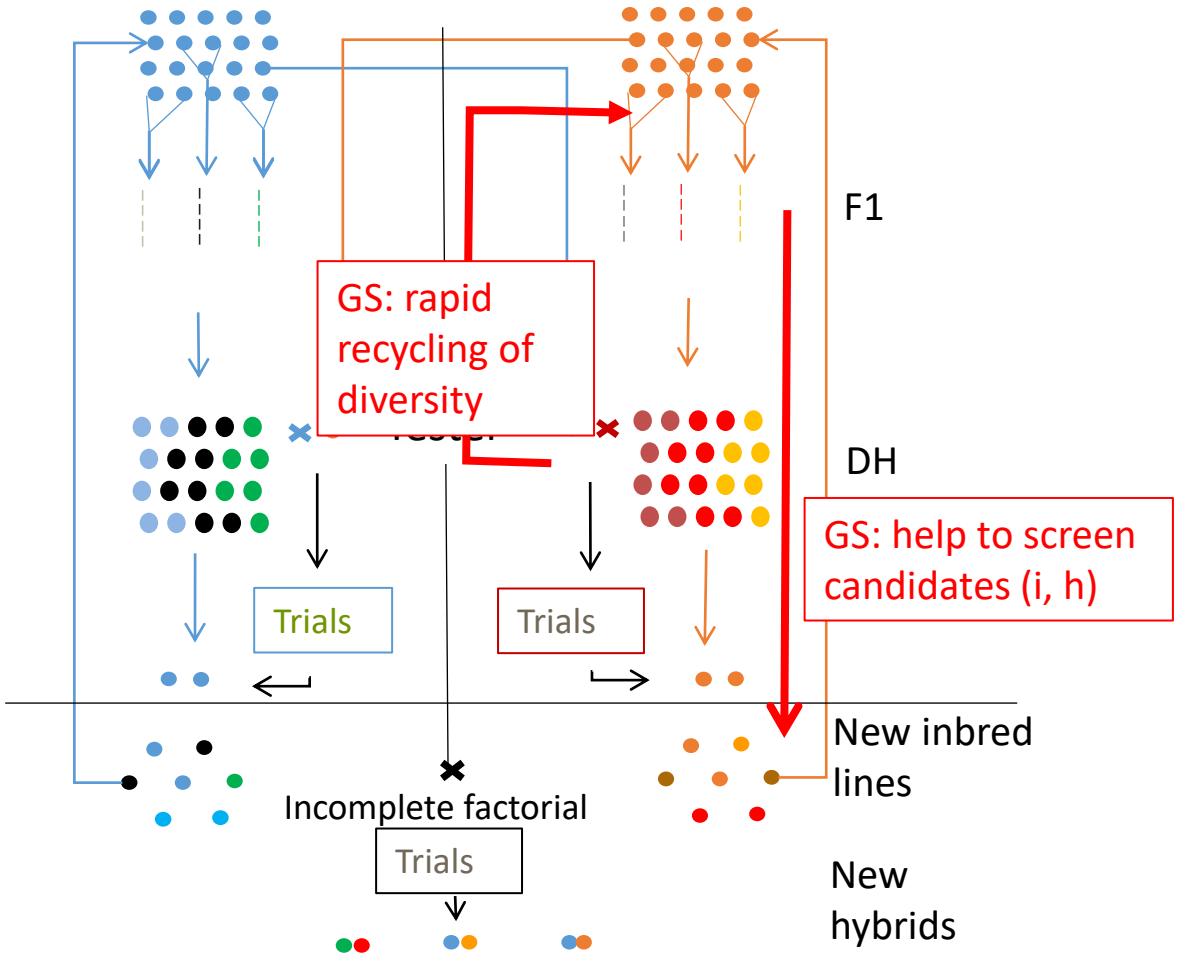


$$\Delta\mu\ /T = i_{(t)} h_{(t)}\ \sigma_{A(t)}/T$$

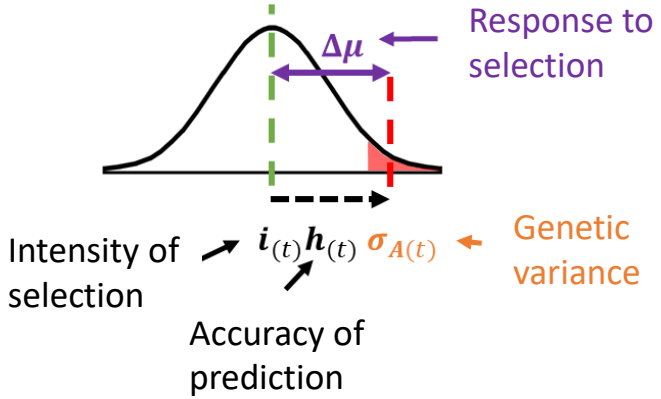


-More global gestion of variability
-Replacement of part (or all) of the phenotyping by genomic prediction
« Sparse testing » (ex: Atanda et al. TAG 2020..)

Use of GS in maize breeding programs

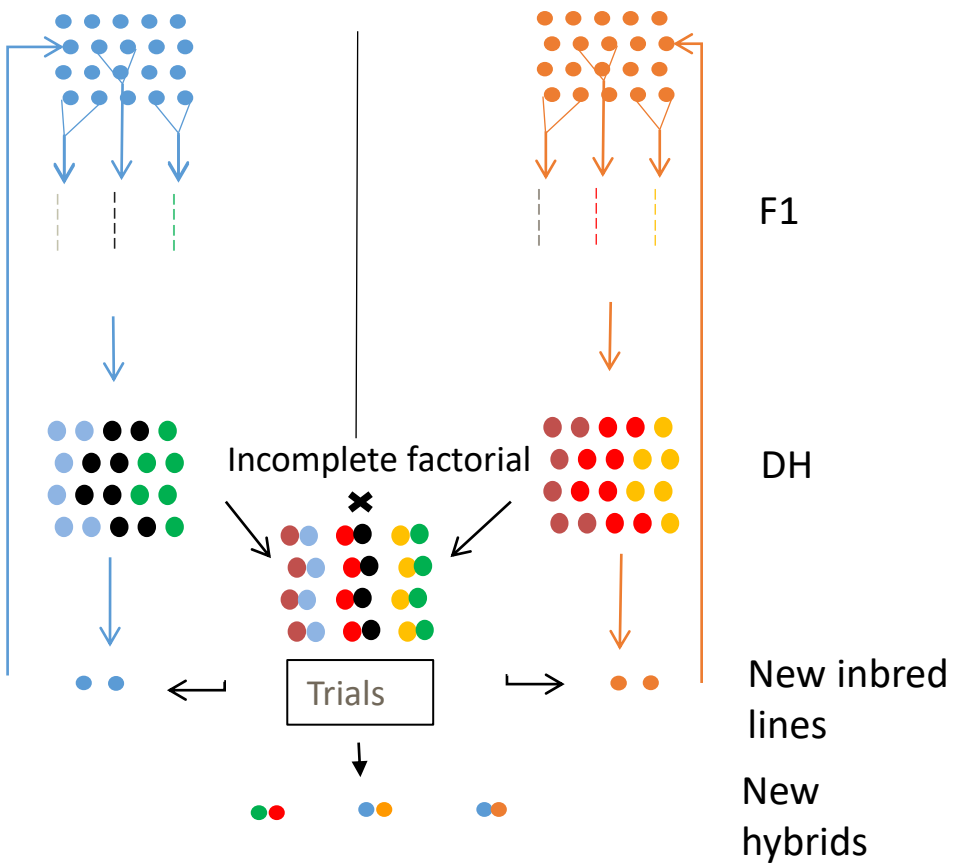


$$\Delta\mu\ /T = i_{(t)} h_{(t)}\ \sigma_{A(t)}/T$$



-More global gestion of variability
-Replacement of part (or all) of the phenotyping by genomic prediction
« Sparse testing » (ex: Atanda et al. TAG 2020..)
-Rapid recycling of diversity (before phenotyping)

Revisit with GS reciprocal selection



Replace tester evaluation by genomic predictions calibrated on a **sparse factorial between unselected candidates**

Main advantages:

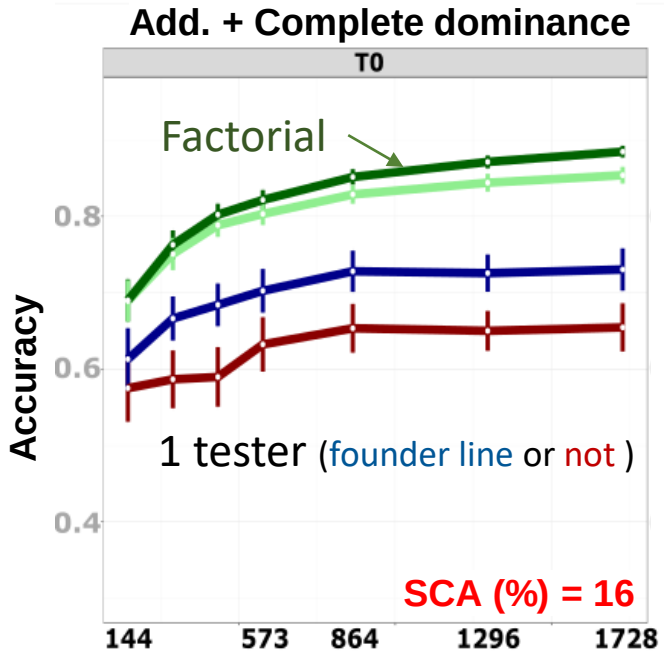
- Avoid phenotyping done independantly in each group
- Increase the number of candidate lines evaluated at a fix number of hybrids
- Potentially select directly on GCA + SCA

Revisit with GS reciprocal selection

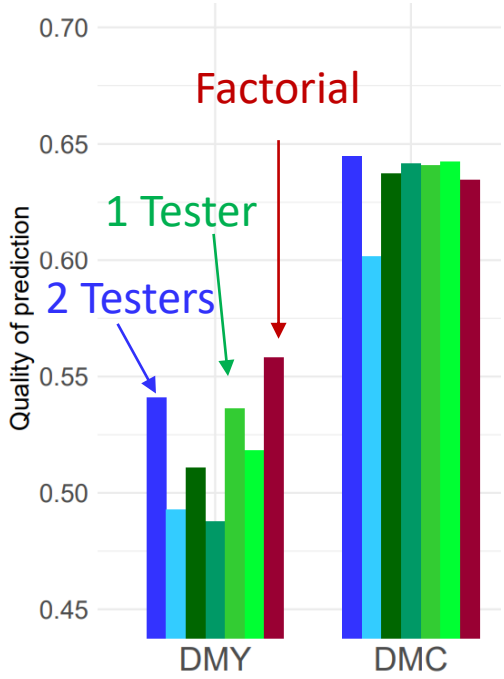
Comparison of tester vs factorial using

PROmaïs SAMMCR project

Simulations (*Seye et al. TAG 2020*)



Experimental results (*Lorenzi et al. TAG, in review*)



Comparaison for calibration set sizes of 180 hybrids

Simulations show an advantage of Factorial designs over Tester designs for traits showing SCA

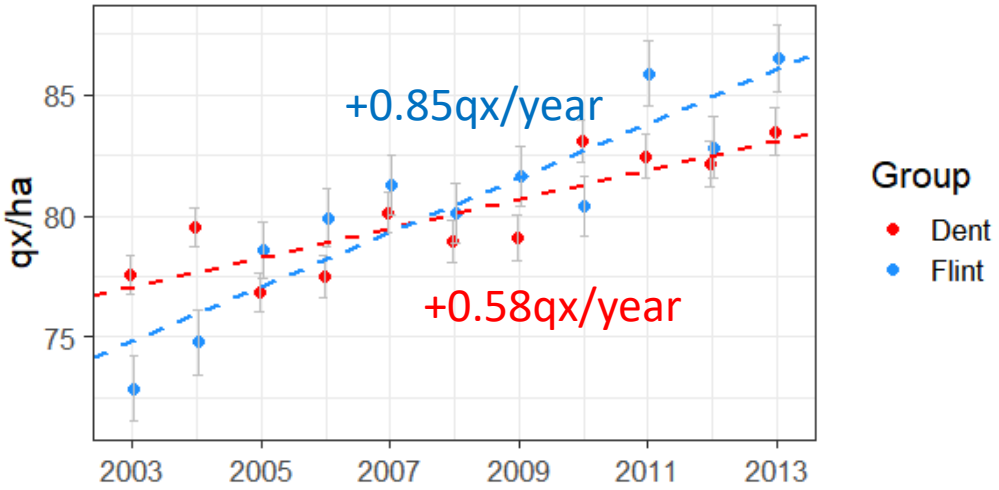
Interest validated experimentaly on average over traits and clearer for yield (DMY)

⇒ Approach wich strongly modifies breeding practices (isolation nursery replaced by hand-made crosses) currently tested by several breeding companies

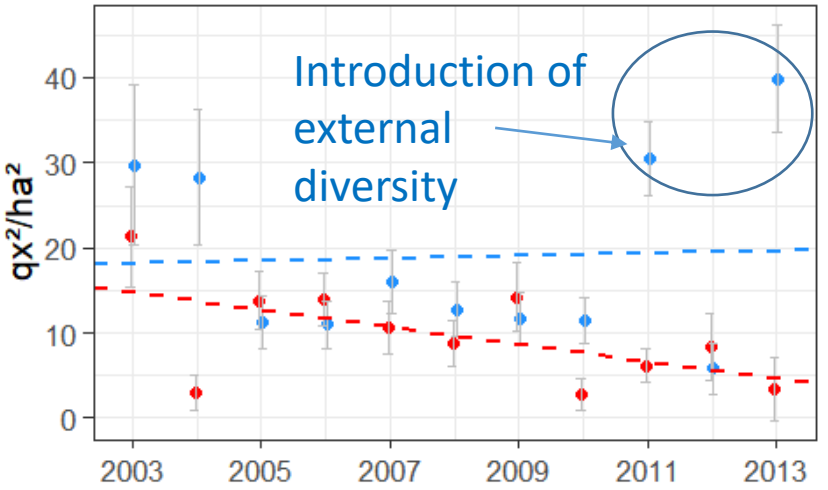
Structuration in heterotic groups increases the risk of diversity loss in breeding programs

Allier et al. TAG, 2019a

Evolution of breeding values

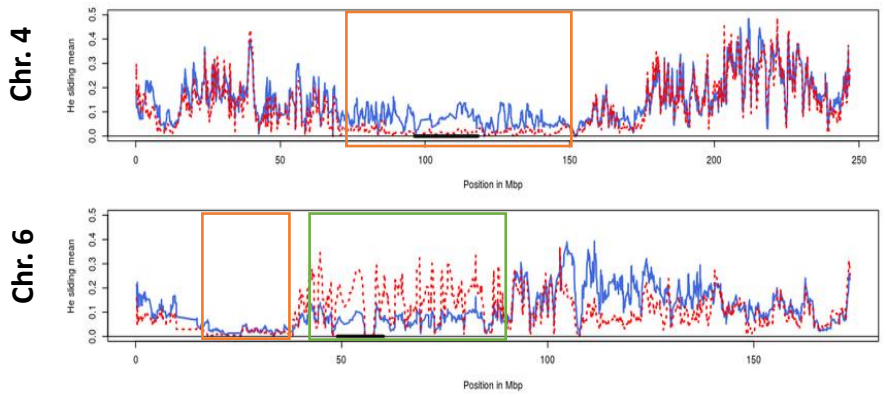


Evolution of genetic variance



- Genetic diversity
  Dent < Flint pool
- Significant reduction in Dent pool

<u>Local genetic diversity</u>
along the genome measured using markers (dent pool)



2003-2009
2010-2016

- Specific patterns of genetic diversity evolution
- Large regions near to fixation (e.g. pericentromere Chr. 4)

=> How to better use diversity in breeding and how to best integrate new sources of diversity ?

Key challenges:

- **Food security** in face of **global climatic** and **societal changes**
- **Constantly rising demands** for plant-based products

→ **More accurate** and **faster breeding** for **new and changing targets**

👍 Predictive Breeding (**genomic selection**) is key to reach this target

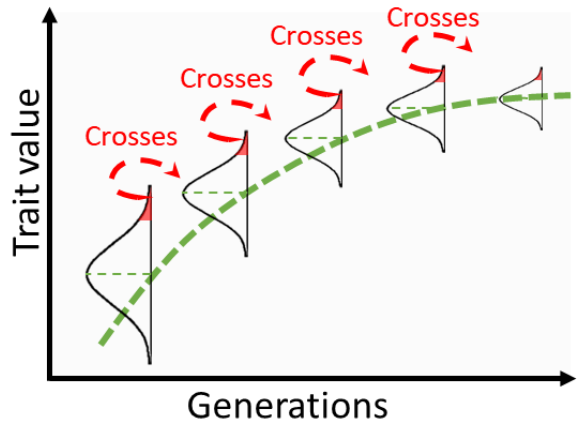
👎 Recurrent truncational genomic selection in closed populations could rapidly **erode genetic diversity** and thus **limiting long term genetic gain** and **breeding sustainability** (Jannink 2010, Rutkoski *et al.* 2015, Allier et al. 2020).

Breeders need to:

- **Efficiently convert causal diversity into genetic gain** while **keeping decent level of still neutral diversity**
- Anticipate the farmer needs and constantly seek for external sources of favorable alleles in diverse **Genetic Resources** (GR) and **efficiently harness** them into breeding

→ Some **GS methodologies** have been suggested to address this challenge !

Crosses in recurrent selection



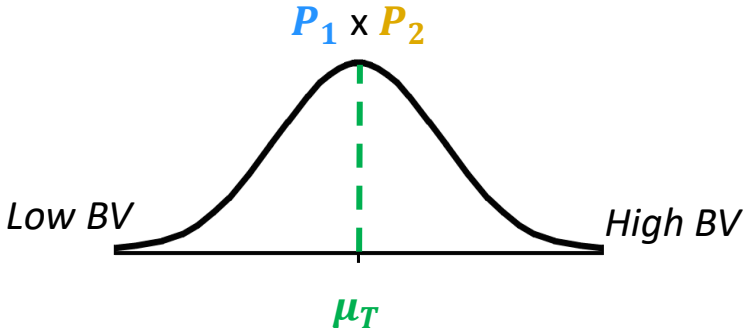
One of the key success factors of recurrent selection is the choice of crosses to generate the next generation

We can select the list of crosses assuming crosses are independent $\begin{pmatrix} P_1 \times P_2 \\ P_1 \times P_3 \\ P_2 \times P_4 \\ P_3 \times P_5 \end{pmatrix}$

→ Truncation selection of crosses

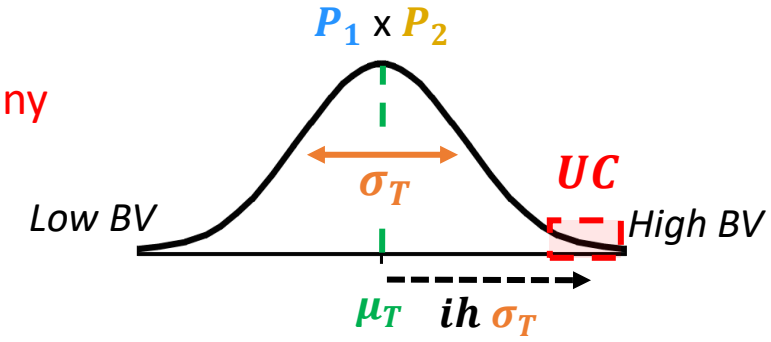
- Maximizing the expected mean breeding value of progeny
  Easily predicted as the mean of parental breeding values

→ **Parental mean (PM)**



- Maximizing the expected mean breeding value of the best fraction of progeny
  Requires the prediction of within family variance σ_T

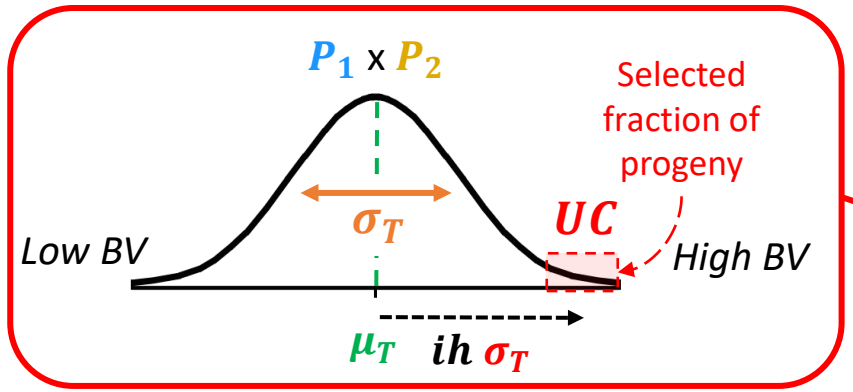
→ **Usefulness Criterion (UC)**



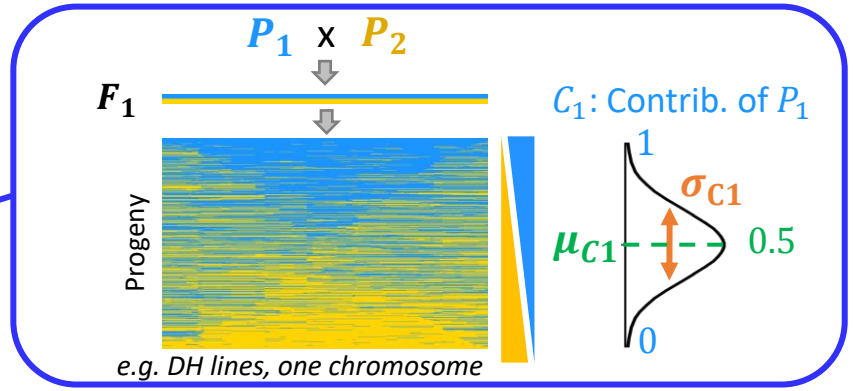
(Schnell and Utz 1975, Lehermeier et al. 2017)

Evaluate crosses based on performance and diversity

Genetic Gain
Usefulness criterion (UC)



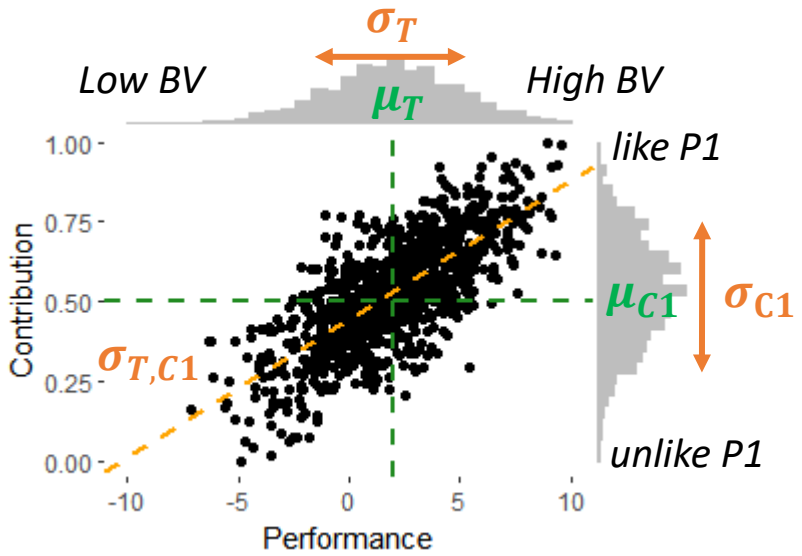
Genome Diversity
Parental contribution (PC)



UCPC

Performance trait(s) and parental genome contribution(s)
as jointly multivariate normally distributed traits

We are able to predict the distribution of performance and parental contributions in progeny …

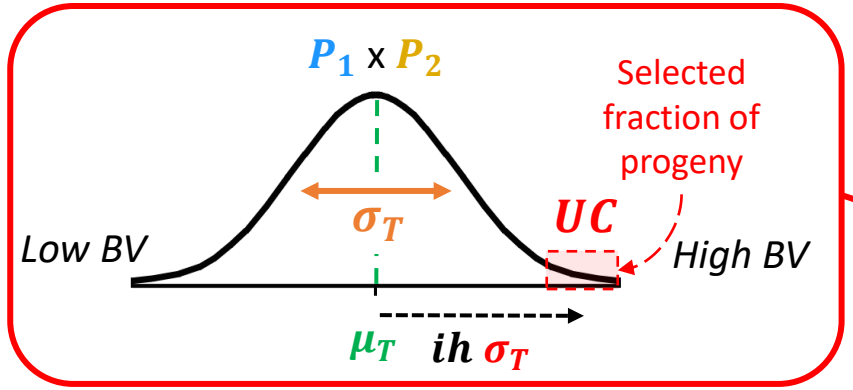


… accounting for **recombination frequency** and **linkage phase** in parents of multi-parental crosses

Extension of Lehermeier et al. (2017)

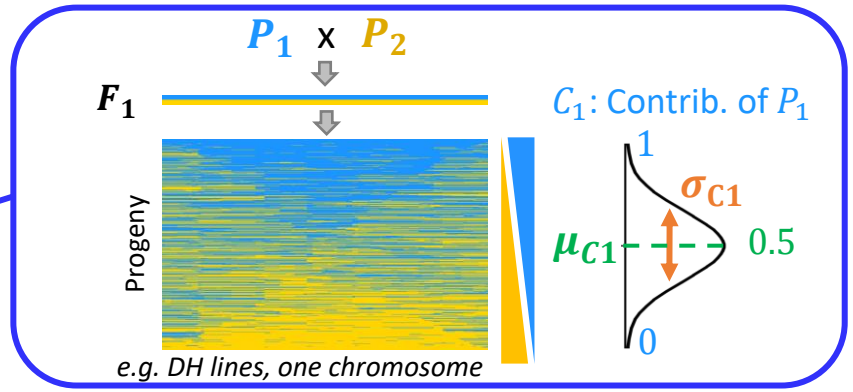
Evaluate crosses based on performance and diversity

Genetic Gain
Usefulness criterion (UC)



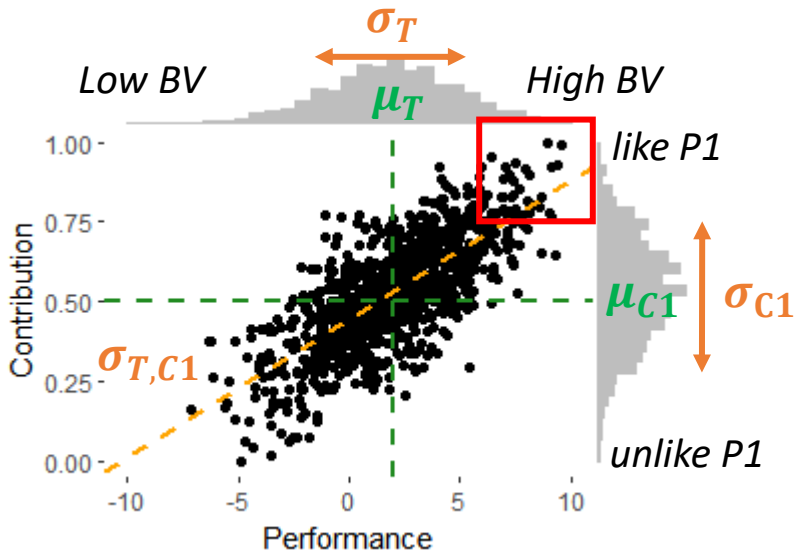
UCPC

Genome Diversity
Parental contribution (PC)



Performance trait(s) and parental genome contribution(s)
as jointly multivariate normally distributed traits

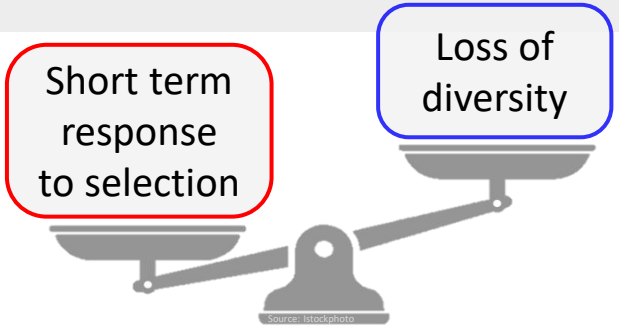
We are able to predict the distribution of performance and parental contributions in progeny …



… accounting for **recombination frequency** and **linkage phase** in parents of multi-parental crosses

Extension of Lehermeier et al. (2017)

Optimization framework



Extension of the Optimal Cross Selection based on the UCPC of crosses

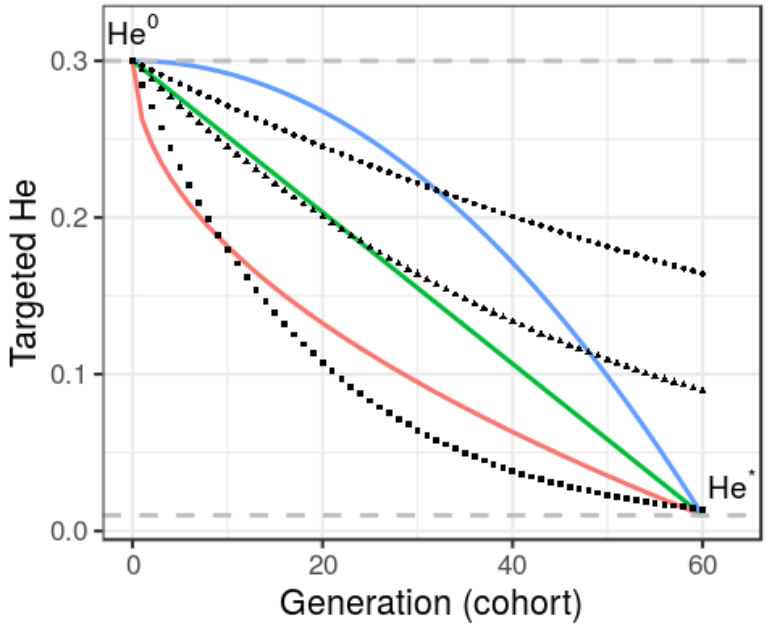
Using a **Differential Evolution** algorithm to find the list of crosses ($\boldsymbol{nc}$) that:
(Storn and Price in 1996)

$$\max_{\mathbf{nc}} V^*(\boldsymbol{nc}) \quad \text{with } D^*(\boldsymbol{nc}) \geq He$$

$V^*(\boldsymbol{nc})$ = expected performance
$D^*(\boldsymbol{nc})$ = expected genomewide diversity (estimated with markers)

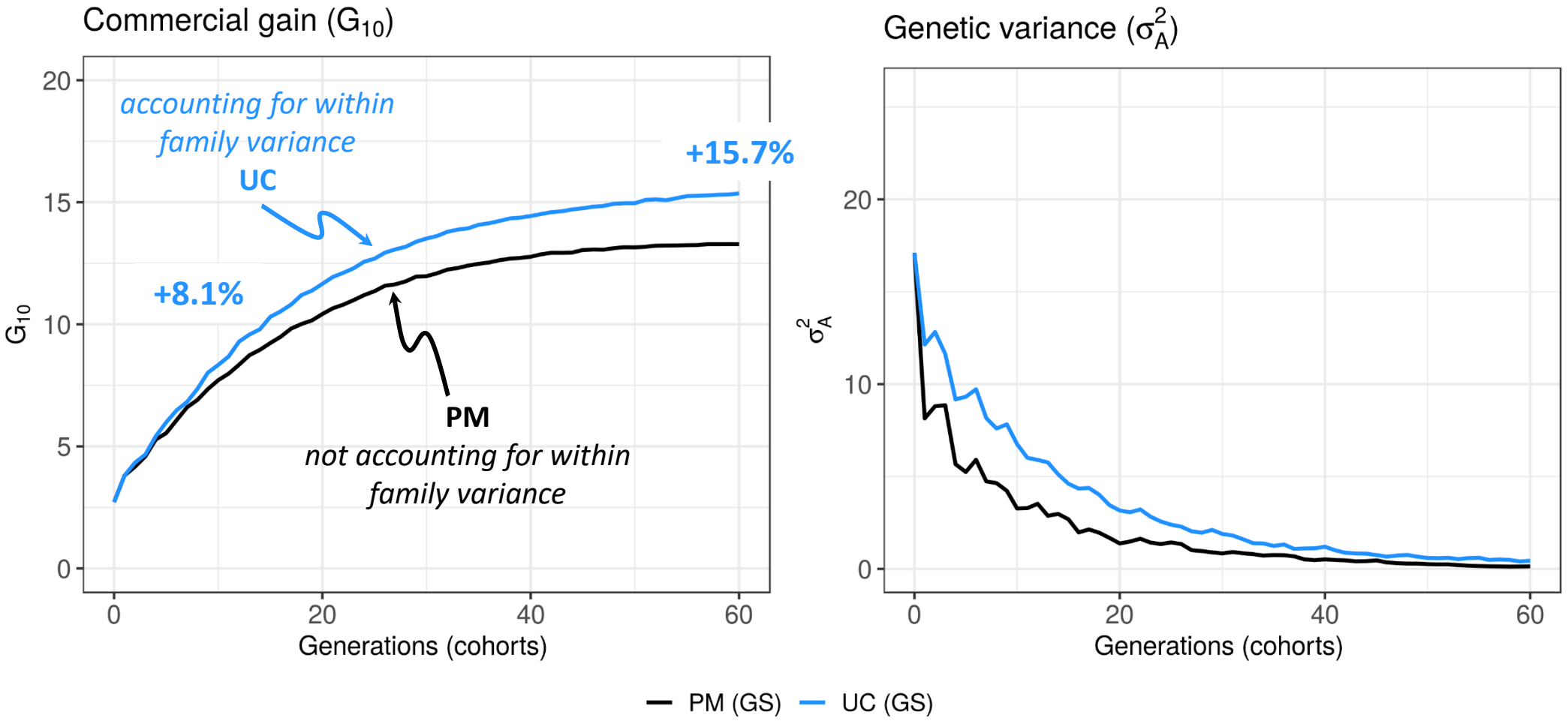
where, He is defined by a **targeted diversity trajectory** (i.e. $He(t), \forall t \in [\![0, t^*]\!]$)

Infinity of trajectories are possible …



… in the following we consider a linear trajectory

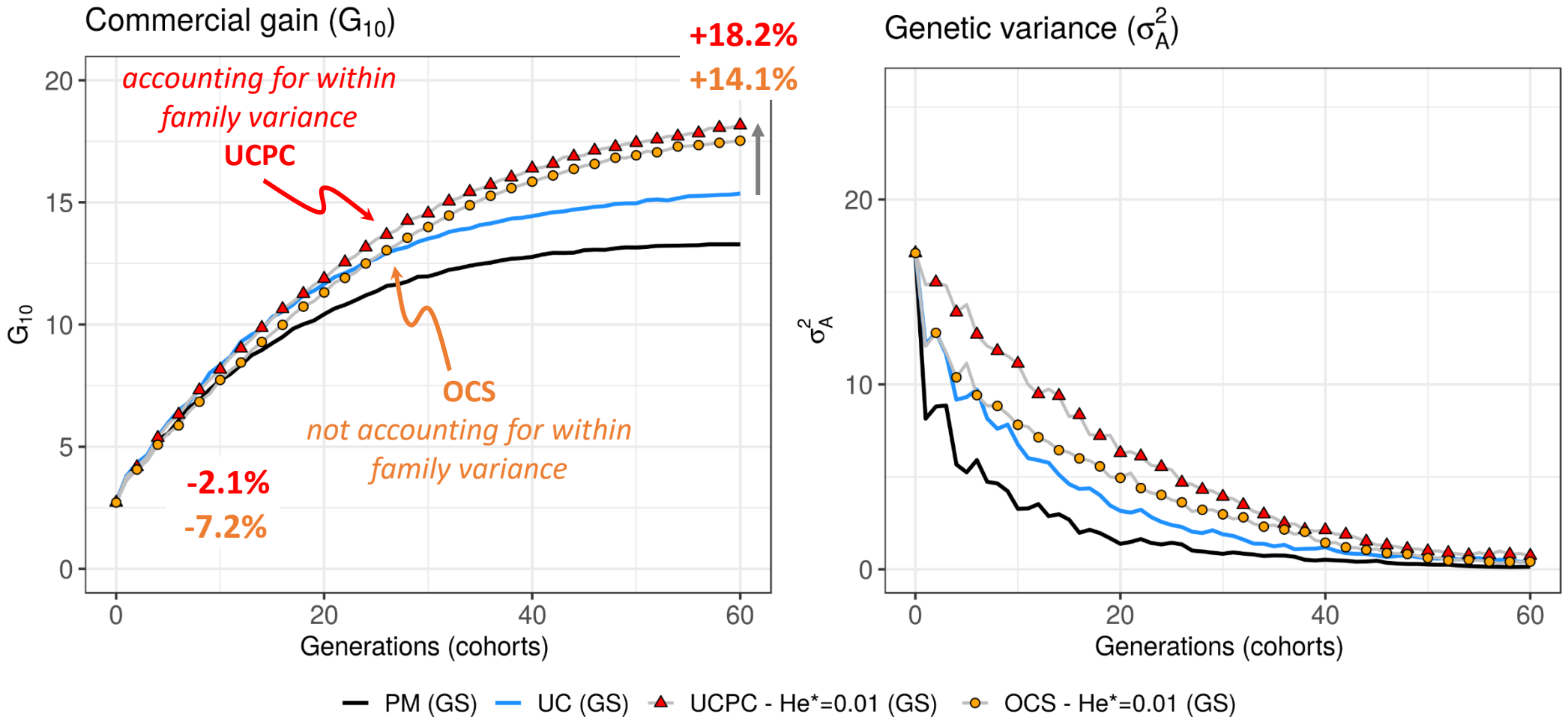
Results



Accounting for within family variance (**UC**) vs. crossing based on parental mean (**PM**):

- Higher short (+8.1%) and long term (+15.7%) commercial genetic gains
- Maintains more additive genetic variance

Results

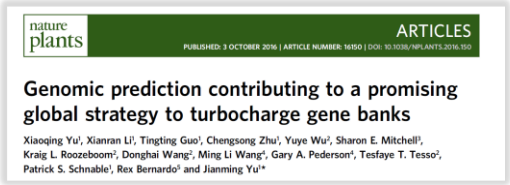


Considering a constraint on diversity (**UCPC** and **OCS** vs. **UC**):

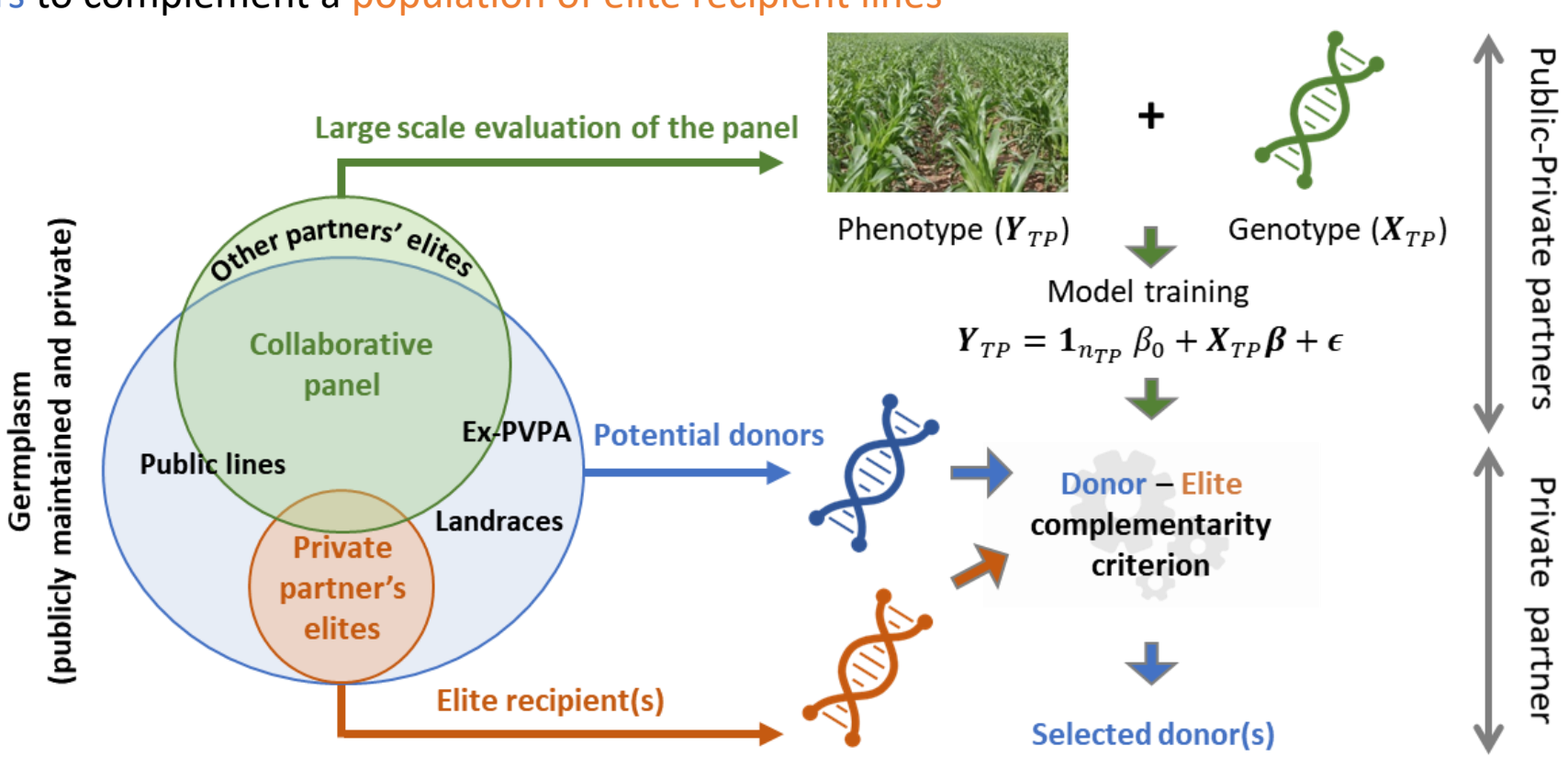
- At short term: limited penalty for **UCPC** (-2.1%) compared to **OCS** (-7.2%)
- At long term: **UCPC** (+18.2%) outperforms **OCS** (+14.1%)

Use of genomic predictions to identify donors to enrich a gene pool

GS for evaluation of Genetic Resources in gene banks as suggested in Yu et *al.* (2016)



GS trained on the Amaizing public-private Dent panel with diverse public lines and proprietary elite lines to identify donors to complement a population of elite recipient lines



Phenotyping in different conditions

Platforms, fields

N-, H20-, Early sowing…

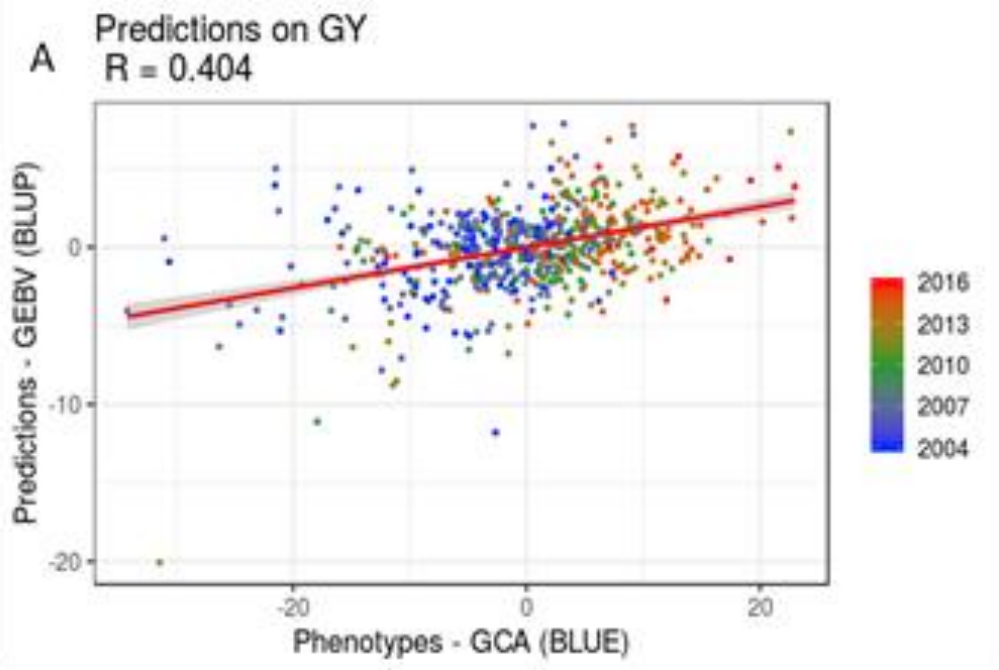
Illustration of the overall approach proposed

Use of genomic predictions to identify donors to enrich a gene pool

Allier et al. TAG 2020

Calibration on a diversity panel including both elite and sources of diversity can predict the genetic gain made in an elite breeding program

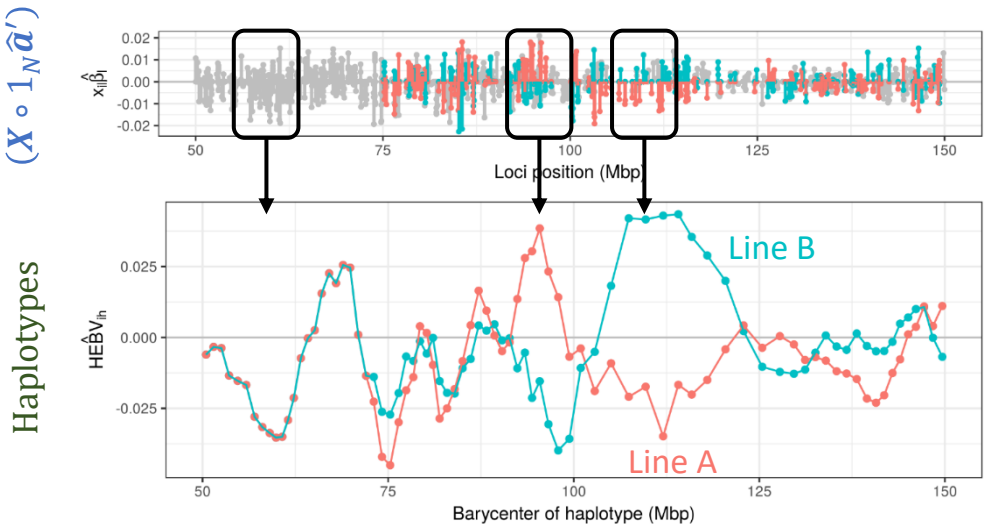
=> Suggests that GS can be used to identify interesting sources of diversity



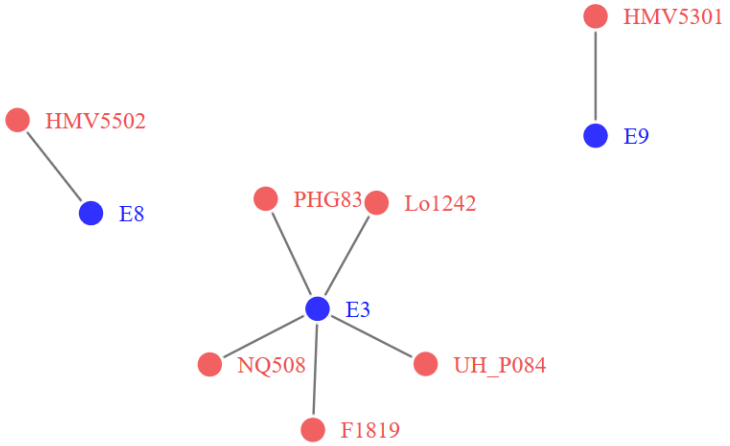
Use of genomic predictions to identify donors to enrich a gene pool

Allier et al. TAG 2020

Use of marker effects ($\widehat{a}$) to define + / - haplotypic blocs



- Allelic effect of donor line A
- Allelic effect of elite line B
- Same allelic effects of A et B



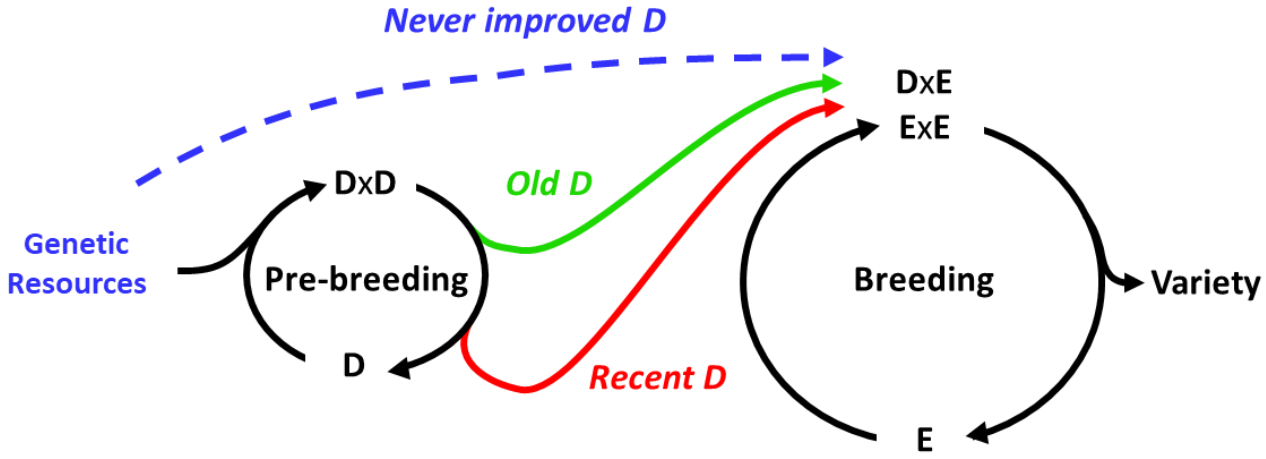
Iterative process to choose the donor lines best complementing an existing pool (here elite lines from RAGT) and how to cross them

Introductions of diversity in elite programs

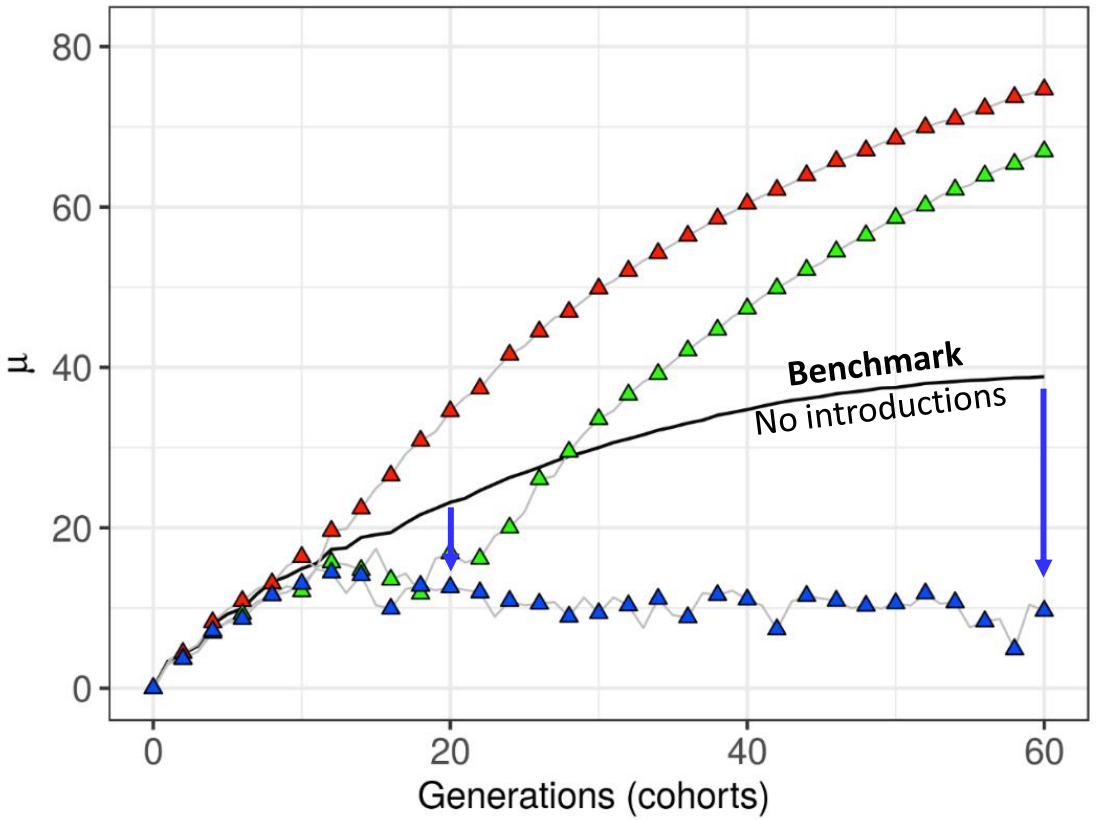
Allier et al BMC Genomics 2020

Scenario:

Open elite program fed directly by **genetic resources**, **old improved resources** or **recent improved resources** to maintain a constant diversity over time (conducted with UCPC)



Genetic Gain without bridging



- **Recent improved D** (5 years old)
- **Older improved D** (20 years old, e.g. "ex-PVP")
- **Genetic Resources** (20 to 80 years old, collections)

→ Direct introductions of improved genetic resources **after pre-breeding** were **favorable at long-term** with possibly a **short-term penalty**

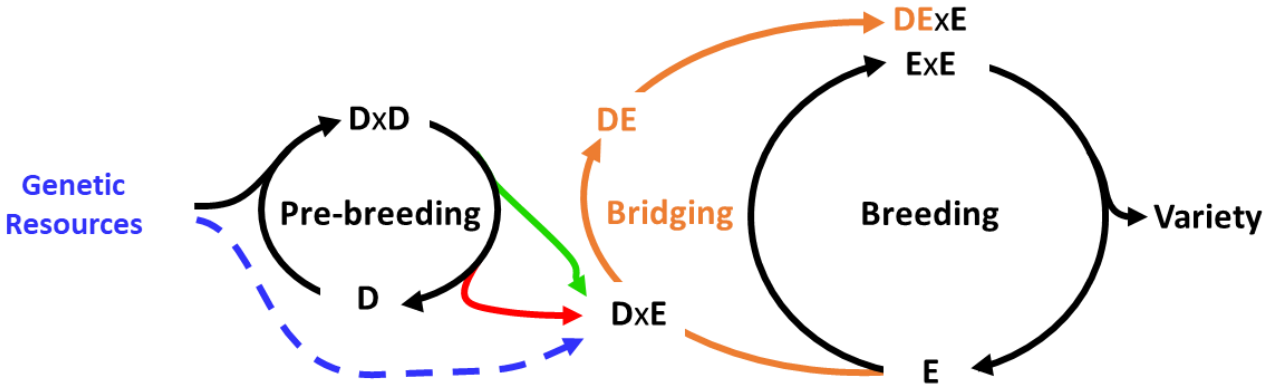
→ Direct introductions from Genetic Resources collections showing high performance gap with Elite are not recommended

Introductions of diversity in elite programs

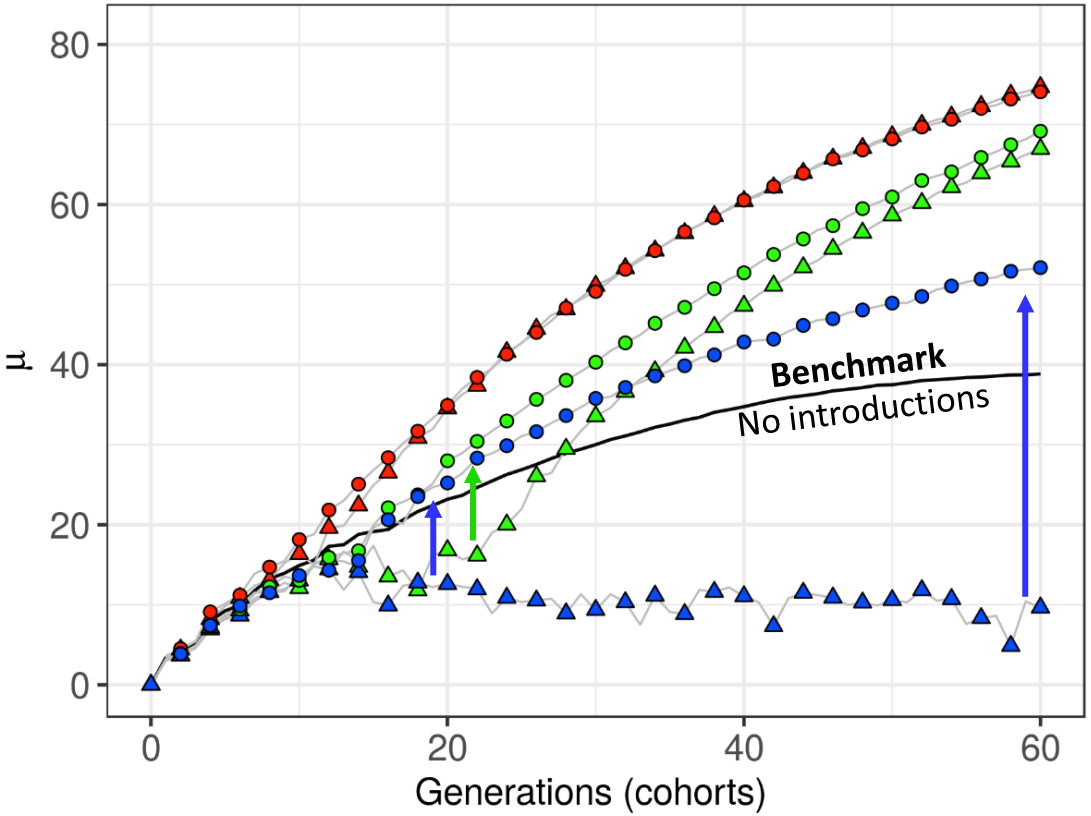
Allier et al BMC Genomics 2020

Scenario:

Open elite program fed indirectly by **genetic resources**, **old improved resources** or **recent improved resources** after a **bridging** to maintain a constant diversity over time (conducted with UCPC)



Genetic Gain with bridging (circles)



- **Recent improved D** (5 years old)
- **Older improved D** (20 years old, e.g. “ex-PVP”)
- **Genetic resources** (20 to 80 years old, collections)

→ Introductions **after bridging reduced short-term penalty** compared to benchmark

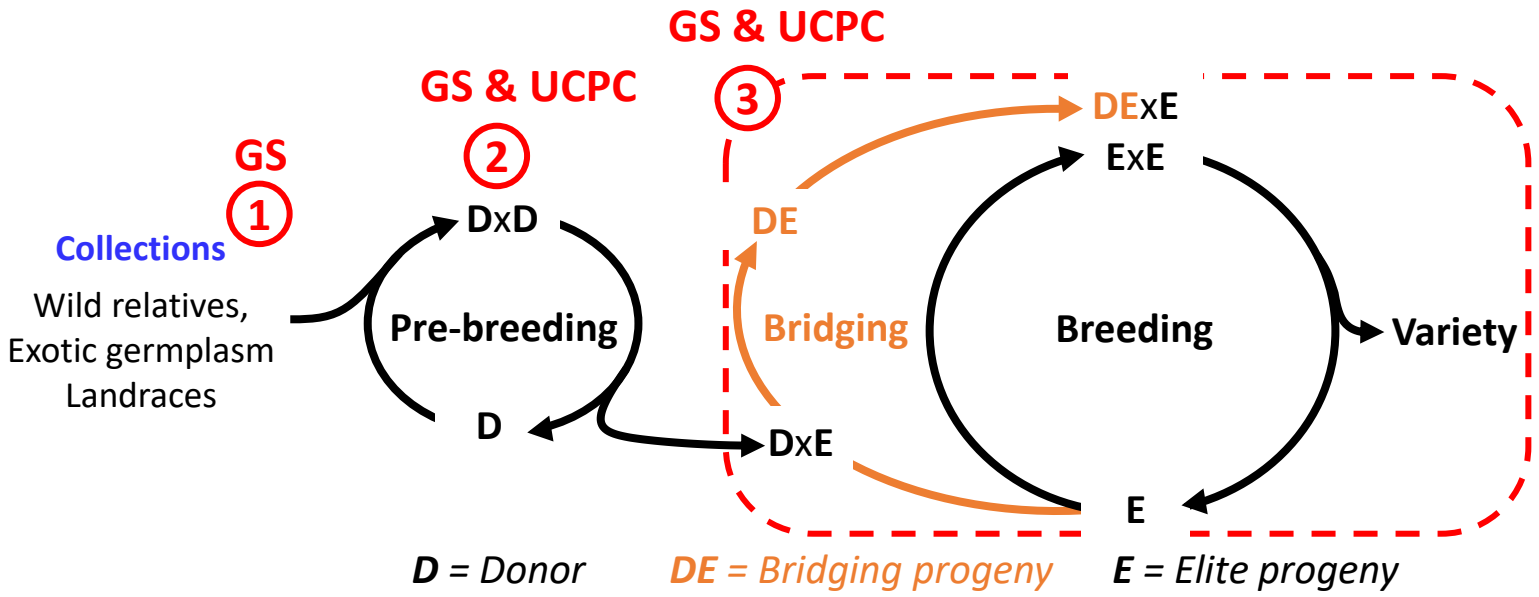
→ Introductions **after bridging from genetic resources collection are beneficial** with limited short-term penalty

Take home messages

GS can be used at different stages of a maize breeding program and to revisit hybrid breeding schemes

GS provides opportunities to better manage diversity in elite material and facilitate the use of genetic resources

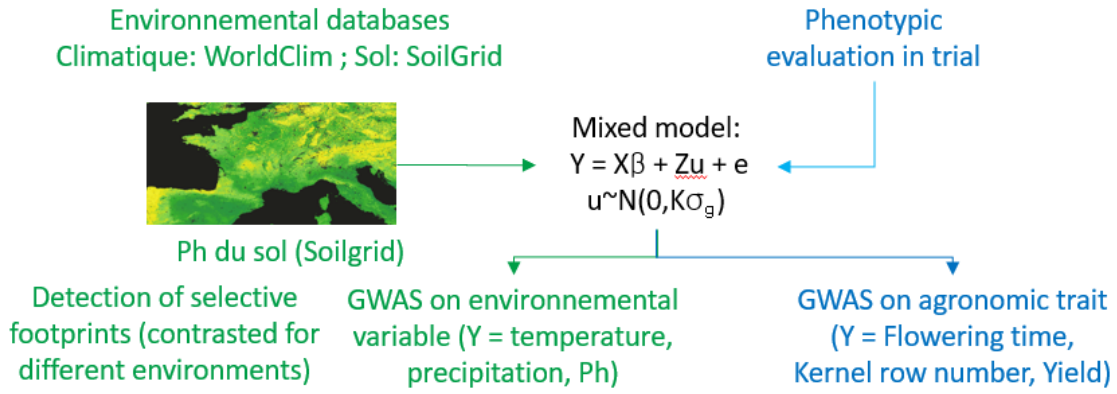
1. **Identification** of Genetic Resources complementing the Elite germplasm assisted by GS
2. **Improvement** of Genetic Resources to reduce performance gap with Elite germplasm while maintaining diversity using GS and optimal cross selection approaches (e.g. Akdemir et al. 2016 in Front. Gen., **UCPC** in Allier et al. 2019 in Front. Gen.)
3. **Introduction** into the Elite germplasm to deliver genetic gain while maintaining diversity constant through recurrent introductions using GS and optimal cross selection approaches



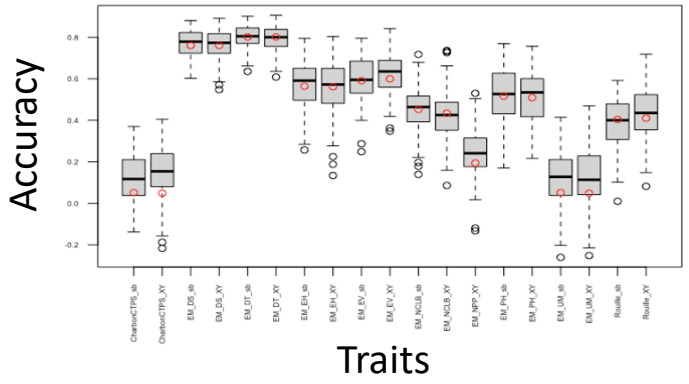
Allier Antoine

Prospects

- Topics related to diversity well adapted to Private-public partnerships. Experimental « participatory » program of introduction of diversity is on going (Promaïs « ValRG », PhD *D. Sanchez*, *A. Charcosset*)
- Need to further characterize genetic resources and more specifically landraces (EVA project, coll. *B. Gouesnard* AGAP)
  - Extract DH lines from Landraces to facilitate their conservation and use in elite germplasm (EHDipop project, *S. BenSadoun*)
  - Genotype landraces on DNA pools (Arca et al., 2021, *S. Nicolas*) and phenotype them to identify original under-used landraces. Perform GS and GWAS to identify adaptative QTLs, predict not yet phenotyped landraces… (PhD *A.Galaretto*)



Preliminary results of inter-landrace GS



Thanks

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

Students:
A. Allier, A. Lorenzi, D. Sanchez, S. Rio, A. Beugnot.,
A. de Walsche, B. Ali,

Collaborators:
GA: Z. Vitezica
UE SMH: C. Palaffre
AGAP: B. Gouesnard
Partners of Amaizing, Promaïs
(SAMMCR, ValRG), Selgen, R2D2

RAGT2n:
C. Lehermeier
S. Teyssèdre

Private partners :
Lidea, Syngenta, Biogemma, Limagrain,
KWS, RAGT, Masseeds, Corteva, Arvalis
Partners of Promaïs and Amaizing projects