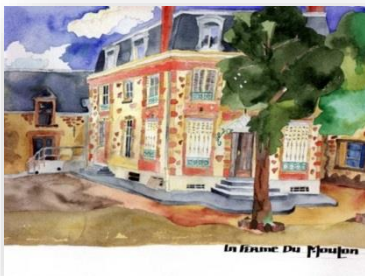# MAKING USE OF NATURAL VARIATION: USE OF GENOMIC SELECTION TO ENRICH THE GENETIC DIVERSITY OF BREEDING PROGRAMS

CropBooster conference, June 11th

A. Allier, S. Rio, L. Moreau, S. Teyssèdre, C. Lehermeier, D. Sanchez, B. Gouesnard, C. Bauland, D. Madur, V. Combes, J. Laborde, C. Palaffre, A. Charcosset

Talk outline

1. Introduction: risk of diversity erosion in a breeding program and impact on selection response

2. Comparing open and closed breeding programs, merits of « bridging » pre-breeding populations to introduce external sources

3. Detection of promising diversity sources: inbred lines cooperative panels and genomic prediction
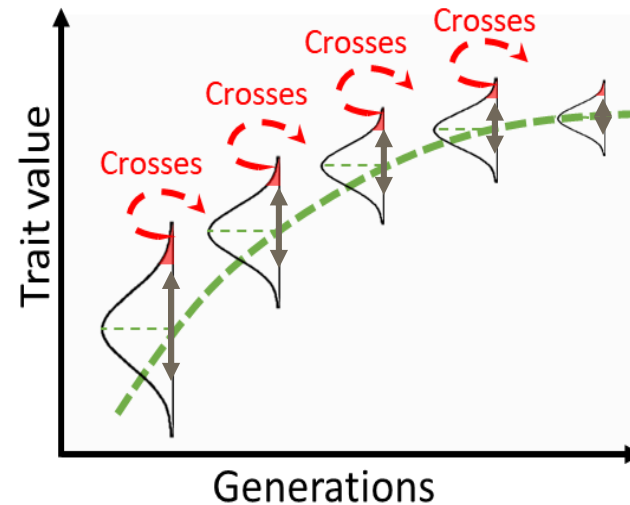
Why is diversity essential?

Breeders' equation for genetic gain
(Lush 1937)

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

Selection
intensity

Selection
accuracy

Genetic variance,
depends on:
Diversity at QTLs
LD between QTLs

Breeding can lead to lose diversity along cycles
Genomic selection can speed up this loss

Need to: **- evaluate its evolution**
            - limit loss while maintaining gain
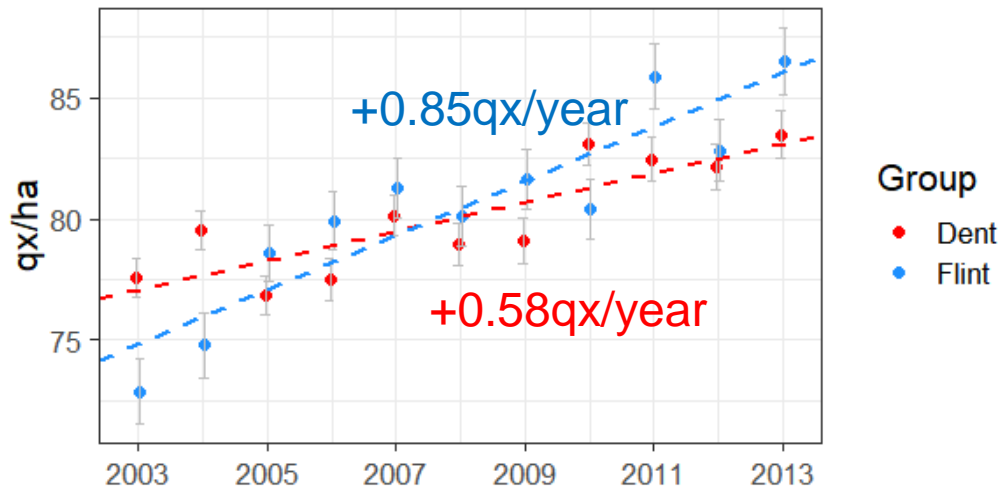   **- search for new diversity sources**

Example: evolution of variance and diversity in RAGT early maize program (Allier et al., 2019)
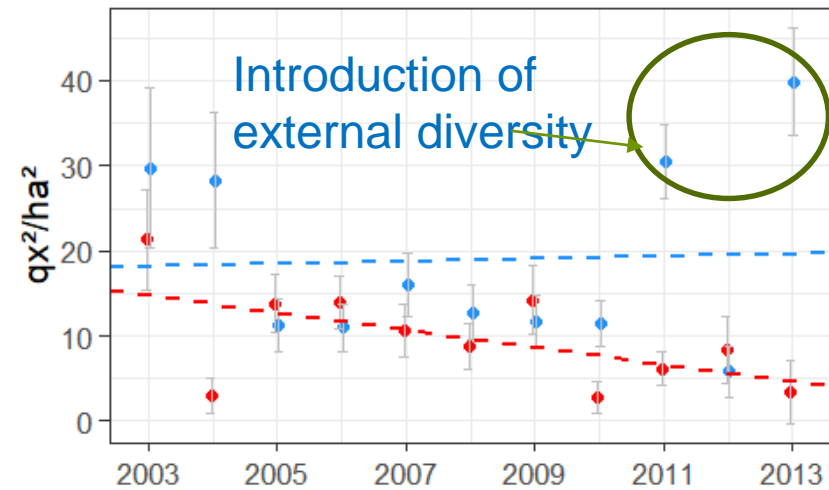
*PhD of A. Allier
(cifre RAGT)*

Two complementary populations (Flint and Dent)

Evolution of genetic values

+0.85qx/year

+0.58qx/year

Evolution of genetic variance

Introduction of external diversity
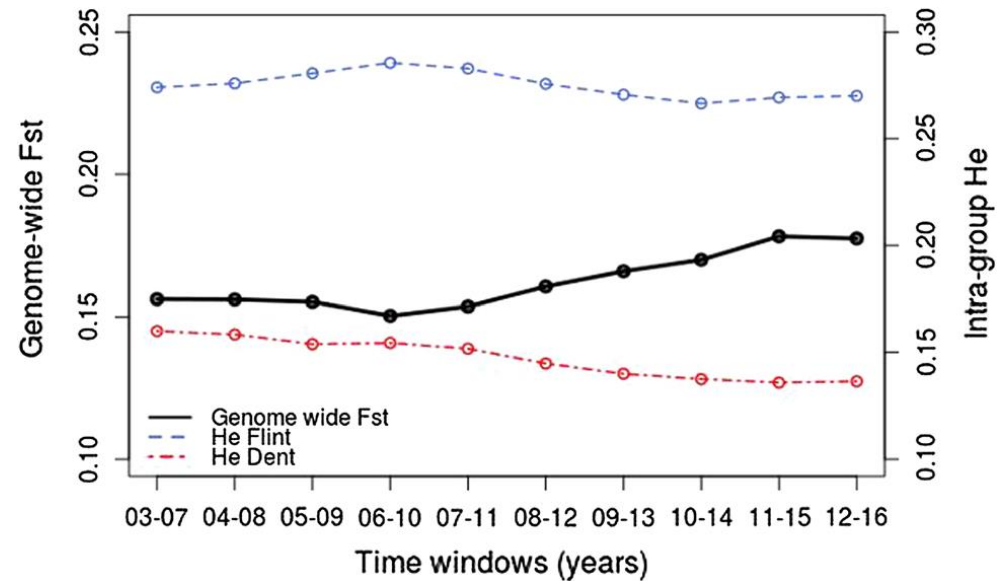
Group
● Dent
● Flint

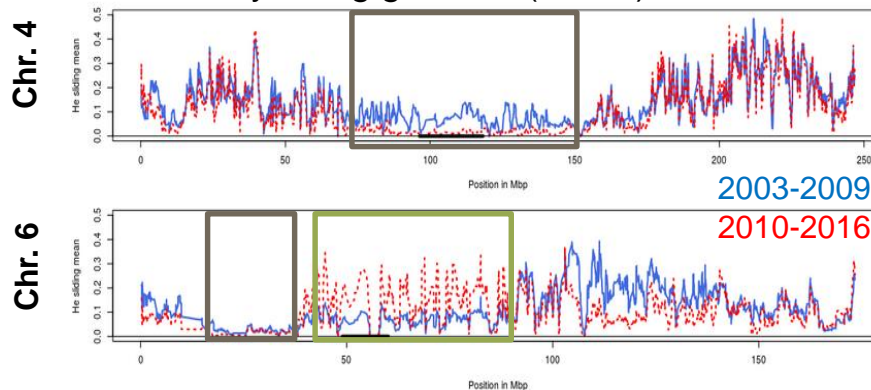*-> lower genetic gain and decrease in variance in the dent group*

At the molecular level

Flint and Dent genetic diversity at neutral markers (He, right axis) and differentiation (Fst, left axis) measured on 5 years sliding windows (Allier et al. 2019)

-> decrease in diversity in the dent group

Genetic diversity along genome (Dents)

2003-2009
2010-2016

-> Some regions are totally fixed now (no diversity left …)

Managing diversity is important to sustain selection efficiency

Which genetic resources to sustain variation (traits previously selected) or enrich it (new target traits)

Age of genetic resources

- Old varieties (e.g. Landraces)
- Inbred lines extracted from these (or from old breeding generations)
- More recent cultivated varieties / parents (hybrid parental lines)

Adapted / non adapted

- ✓ To environment and uses
- ✓ To organization of the program in heterotic groups (difficulty for using commercial hybrids)

-> How to identify relevant resources for given breeding objectives?

-> Which breeding methods to use them?

Talk outline

1. Introduction: risk of diversity erosion in a breeding program and impact on selection response

2. Comparing open and closed breeding programs, merits of « bridging » pre-breeding populations to introduce external sources

3. Detection of promising diversity sources: inbred lines cooperative panels and genomic prediction

Simulated populations

RESEARCH ARTICLE   **Open Access**

## Optimized breeding strategies to harness genetic resources with different performance levels

Antoine Allier[1,2*] , Simon Teyssèdre[2], Christina Lehermeier[2], Laurence Moreau[1] and Alain Charcosset[1*]

Simulation model (RAGT Simulator)

- Breeding founders / Genetic resources: true Genotypic data (338, Amaizing Dent Panel)
- Architecture: 1,000 additive QTLs with $\boldsymbol{\beta} \sim N(\mathbf{0}, 0.05\,\boldsymbol{I})$
- + 2,000 neutral SNPs
- Heritability: 0.73 in founders (*1 year, 4 locations, no GxE*)

Closed commercial breeding program

- **20 years of burn-in**:

Phenotypic selection (PS) from 10 founders (/ 57 Iodent)

- **60 years post burn-in**:

Genomic selection (GS), management of diversity decrease
Estimated marker effects $\widehat{\boldsymbol{\beta}}$ updated every year
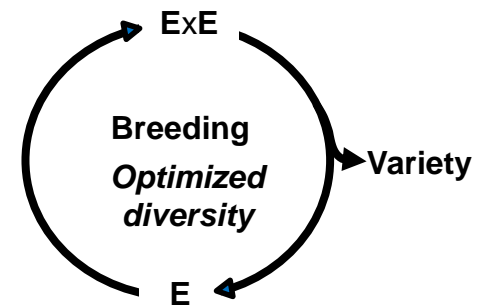
## External (pre-)breeding program (-> improved diversity sources)

- Same process from 40 founders (all panel)
- Can be viewed as pre-breeding pool or competitors programs
- -> Time constraint in accessing the outcomes of this program :
  - ✓ 5 years (mimics UPOV convention, for autogamous species)
  - ✓ 20 years (mimics US PVP for hybrids)

Direct introductions without bridging

Open elite program fed directly by external resources to maintain a constant diversity over time (conducted with "UCPC")

D = Donor
E = Elite progeny
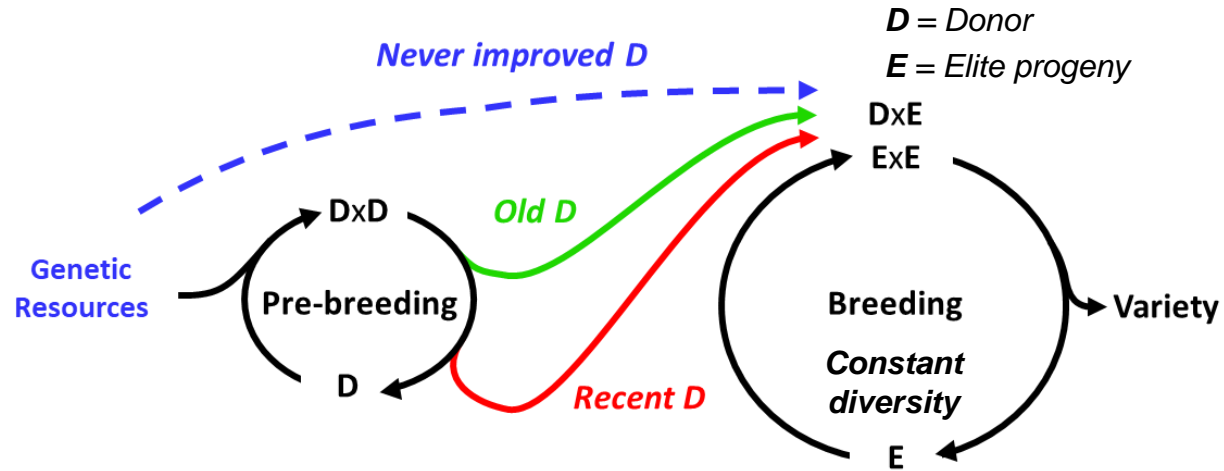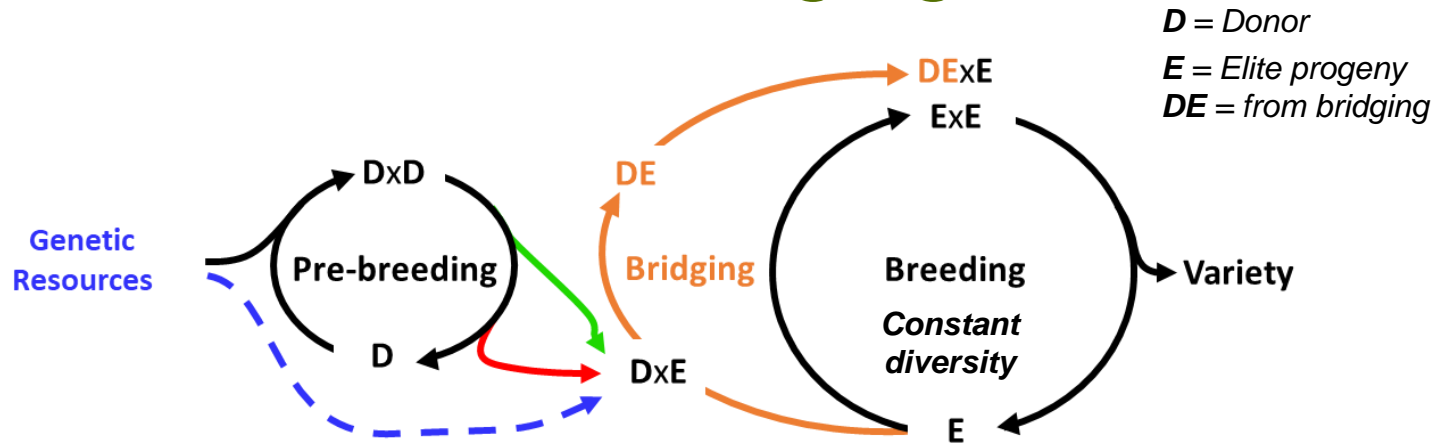

Genetic gain with introductions or not
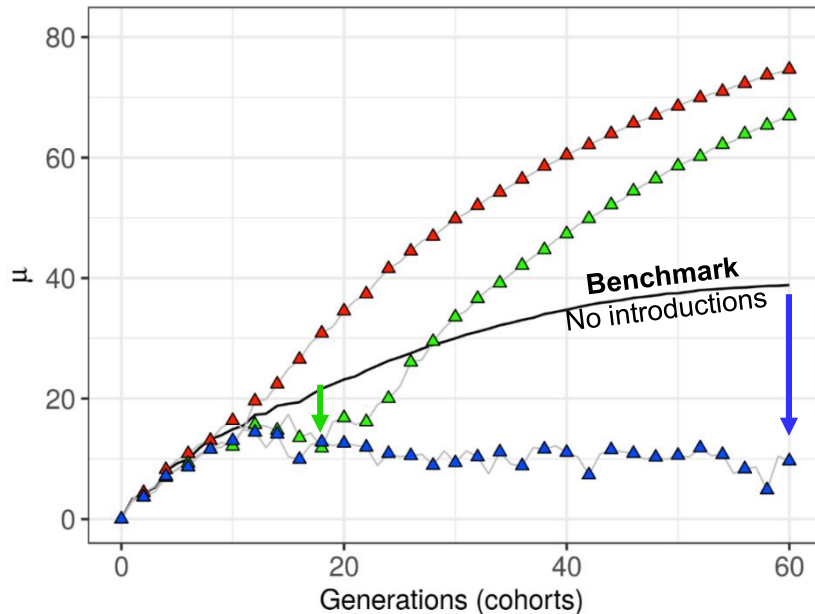
▲ **Recent improved D** (5 years old)
▲ **Older improved D** (20 years old, e.g. "ex-PVP")
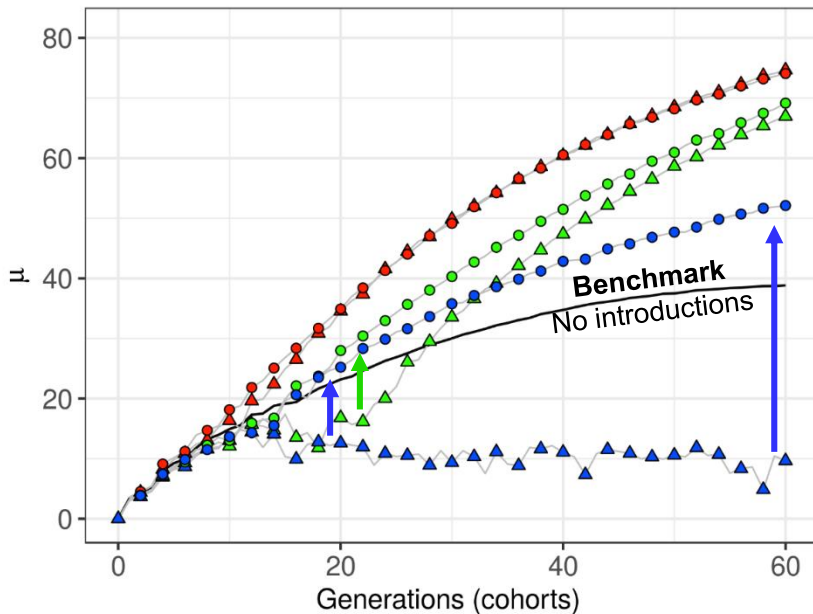▲ **Genetic Resources** (20 to 80 years old, collections)

✓ *With the exception of older GenRes, introductions are beneficial / Benchmark*
✓ *Possibly at the cost of short term gain*

Indirect introductions after bridging

D = *Donor*
E = *Elite progeny*
DE = *from bridging*

Rk. Same total means as for direct introduction

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)

✓ *Even older GenREs introductions are beneficial*

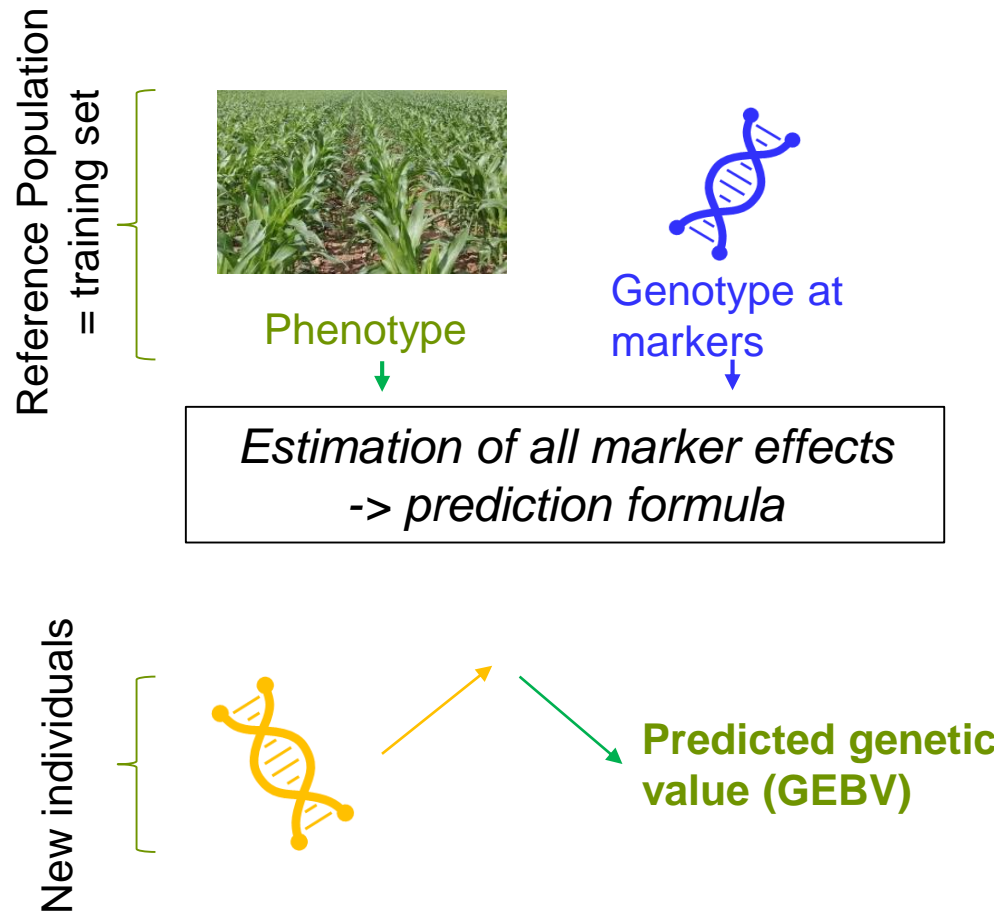✓ *Short term gain closer to Benchmark*

Talk outline

1. Introduction: risk of diversity erosion in a breeding program and impact on selection response

2. Comparing open and closed breeding programs, merits of « bridging » pre-breeding populations to introduce external sources

3. **Detection of promising diversity sources: inbred lines cooperative panels and genomic prediction**

# Genomic prediction aided evaluation of genetic resources for quantitative traits of interest

Reference Population = training set

Phenotype

Genotype at markers

Estimation of all marker effects -> prediction formula

New individuals

Predicted genetic value (GEBV)

Can be applied to genetic resource collections

Genomic prediction in a cooperative maize panel (Rio et al., 2019, TAG)

PCoA on genetic distances

axis number 1 : 11 % of the variance
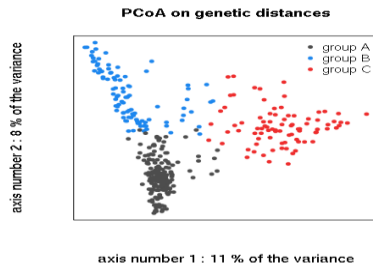
350 dent "public lines" from different generations, including old lines, "ex PVPs" + 49 private recent dent elite lines (7x7 partners),

Leave-One-Out Cross-Validations: each line predicted (GEBV) by all others, comparison to observed value (LSMean)

Yield Index


r=0.73

Male Flowering

r=0.75
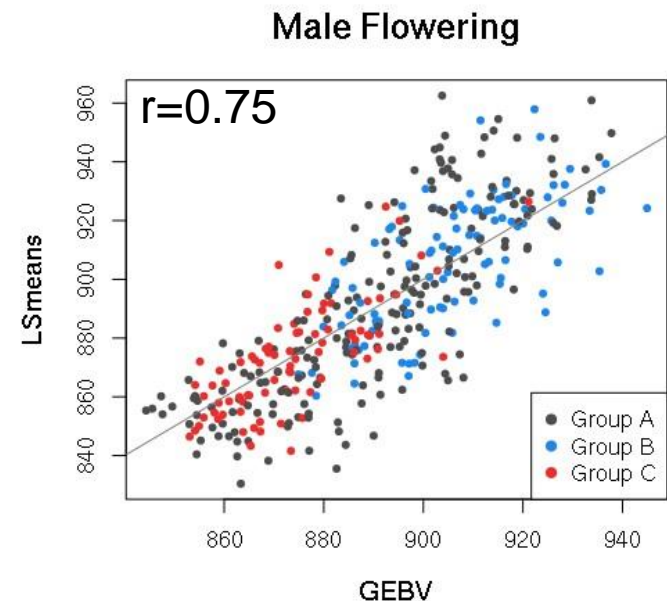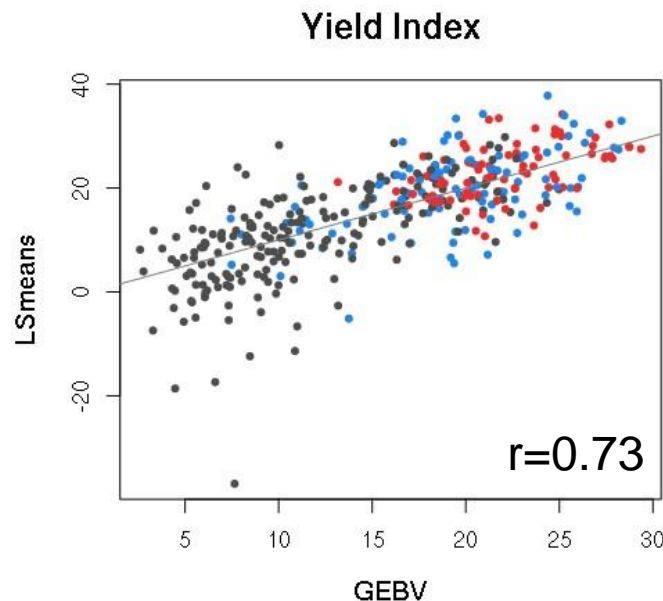
→ Population structure contributes to prediction efficiency, but clear efficiency is also observed within groups

Evaluation of model predictive ability in RAGT material (Allier et al., 2020)

Predictive ability in RAGT Dent material (anti Flint): BLUEs of 594 lines released from 2004 to 2016

| Trait	h	All 13 years	1 year min to max (mean)
Grain Yield	0.347	0.404	-0.062 to 0.722 (0.305)
Male Flowering	0.519	0.495	0.222 to 0.715 (0.476)
Grain Moisture	0.681	0.550	0.286 to 0.811 (0.560)

Predictions on GY
R = 0.404
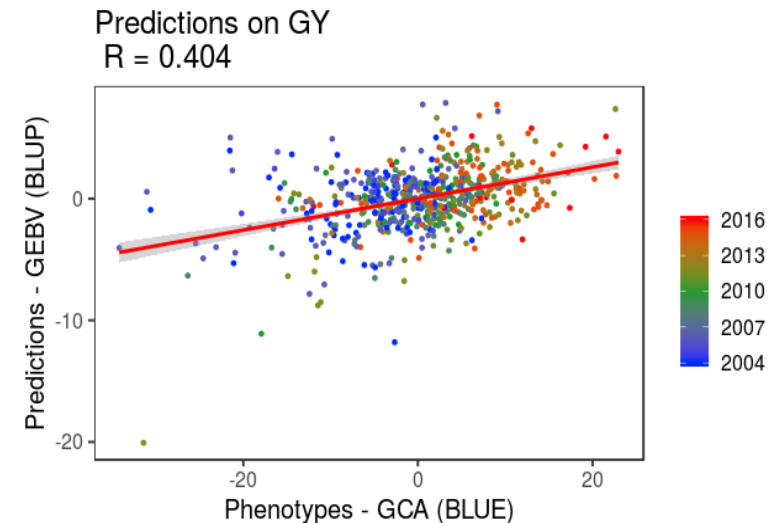
✓ Amaizing Dent Panel has a non null predictive value on elite material for all traits including yield, partly by predicting genetic gain, but not only (rk. Lower intra year values observed in KWS and Euralis germplasm)

✓ *Supports some stability in estimated effects*

Identification of genetic resources

Principle: use of marker effects estimated in a cooperative panel to identify a donor to complement an elite, using *a haplotypic criterion (HEBV)*

$(X \circ 1_N \widehat{\beta}')$ *Matrix of individual loci estimated values [n x m]*

$X: genotypes, \qquad \widehat{\beta}: locus\ effects$
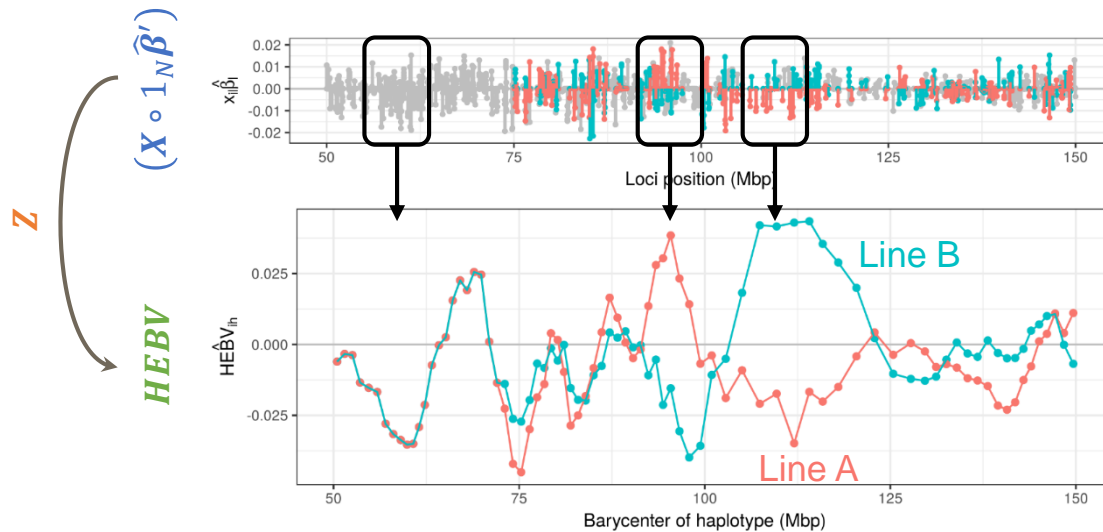
- Effect of Line A allele
- Effect of Line B allele
- Effect of shared allele

$HEBV = (X \circ 1_N \widehat{\beta}')\ Z$

Matrix of local haplotype estimated values [n x nH]

*Design matrix to sum over loci in the same haplotype [m x nH]*

Note. In case of non overlapping haplotypes it is similar to the OHV by Daetwyler et al. 2015

Identification of genetic resources

Principle: use of marker effects estimated in a cooperative panel to identify a donor to complement an elite, using *a haplotypic criterion (HEBV)*

$\left(\boldsymbol{X} \circ 1_N \widehat{\boldsymbol{\beta}}'\right)$ *Matrix of individual loci estimated values [n x m]*

$\boldsymbol{X}: \boldsymbol{genotypes}, \qquad \widehat{\boldsymbol{\beta}}: \boldsymbol{locus\ effects}$
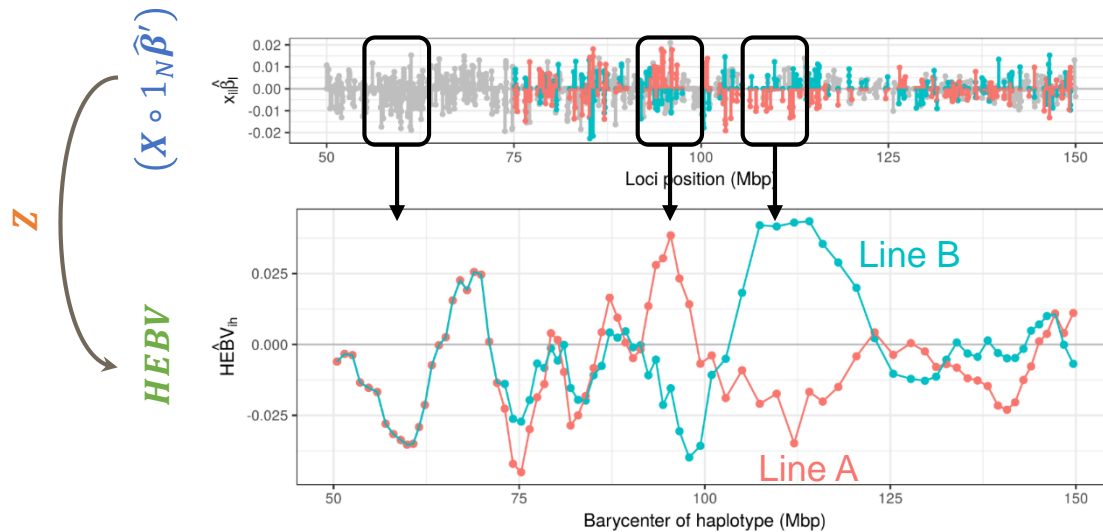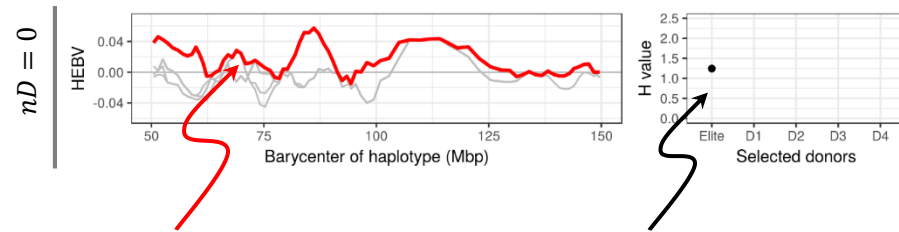

- Effect of Line A allele
- Effect of Line B allele
- Effect of shared allele

Supports here the introgression of 2nd region from line A into line B background

Stepwise selection of donors

Assume we have:

- A population $PopE$ of elites to be enriched (grey curves)
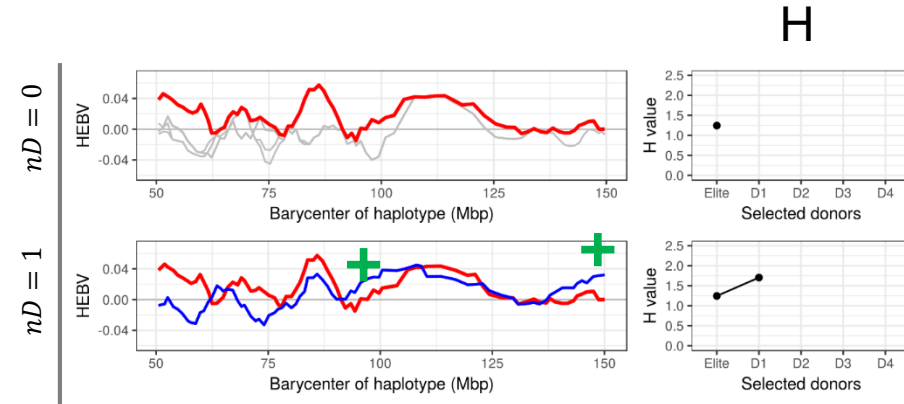

The ideal haplotypic profile within elites
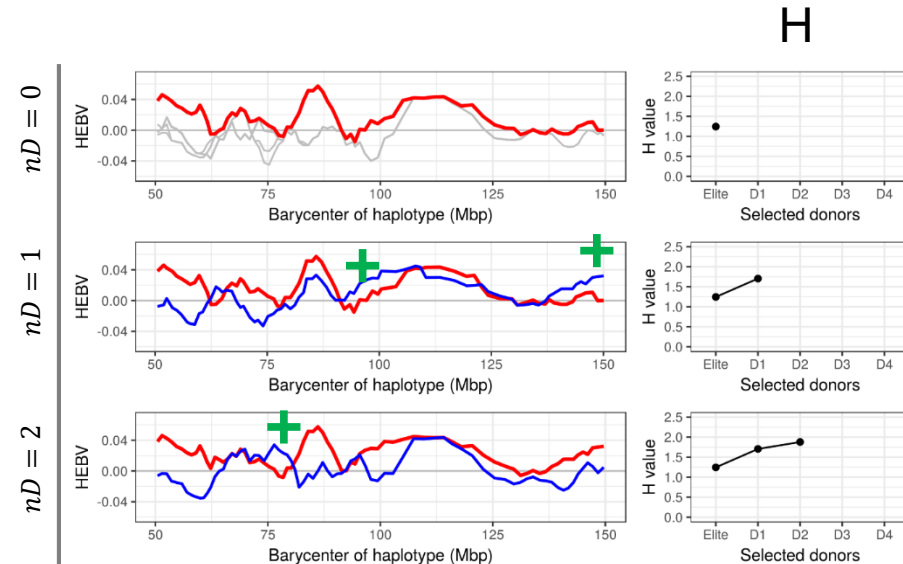
H is the sum of HEBV over the genome

Stepwise selection of donors

Assume we have:

- A population $PopE$ of elites to be enriched (grey curves)
- **A population $PopD$ of candidate donors to select in**

We select the donor that brings most favorable haplotypes (✚) outperforming the ideal profile

H

Stepwise selection of donors

Assume we have:

- A population $PopE$ of elites to be enriched (grey curves)
- **A population $PopD$ of candidate donors to select in**

We select the donor that brings most favorable haplotypes (**+**) outperforming the ideal profile

We update the ideal profile and select the donor that brings most favorable haplotypes (**+**) outperforming the ideal profile

H

Stepwise selection of donors

Assume we have:

- A population $PopE$ of elites to be enriched (grey curves)
- **A population $PopD$ of candidate donors to select in**

We select the donor that brings most favorable haplotypes (✚) outperforming the ideal profile

We update the ideal profile and select the donor that brings most favorable haplotypes (✚) outperforming the ideal profile
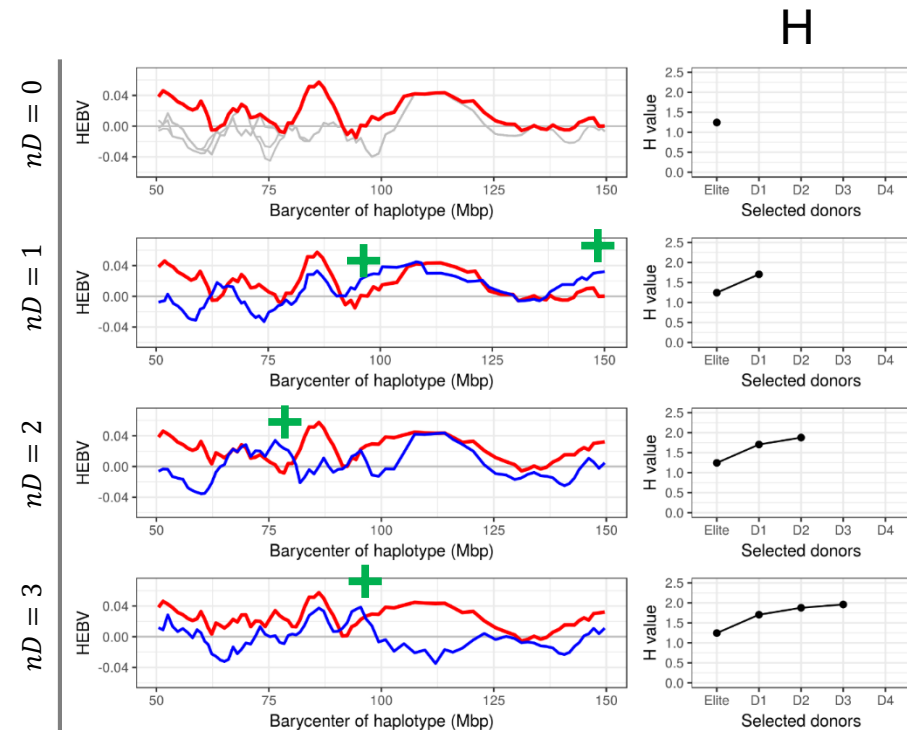

H

Stepwise selection of donors

Assume we have:

- A population $PopE$ of elites to be enriched (grey curves)
- **A population $PopD$ of candidate donors to select in**

We select the donor that brings most favorable haplotypes (**+**) outperforming the ideal profile

We update the ideal profile and select the donor that brings most favorable haplotypes (**+**) outperforming the ideal profile
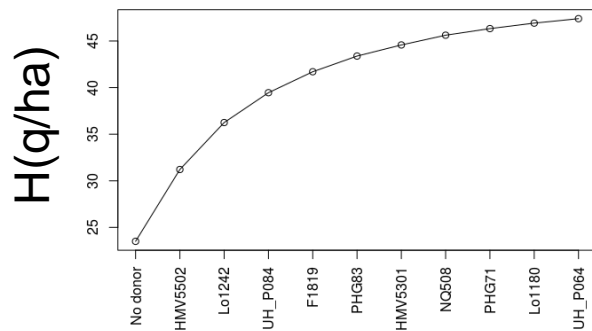
And so forth as a greedy algorithm …

… until there is no more additional gain to consider another donor
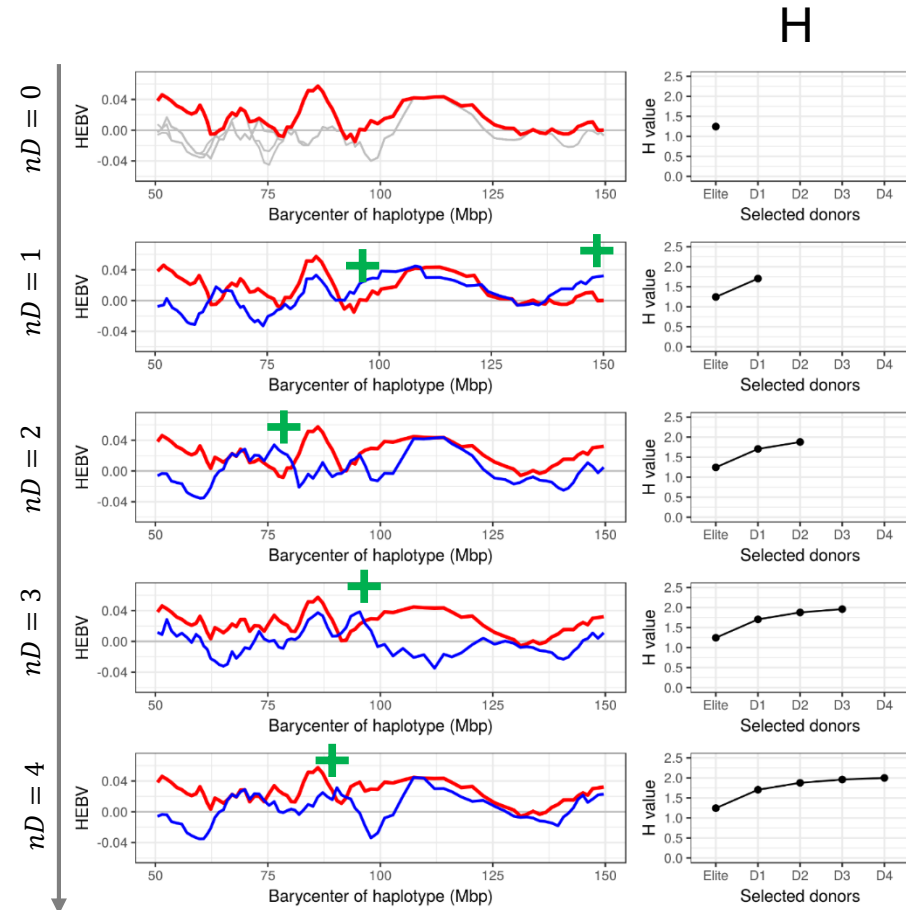
Application case

Application material:

Elites = 10 Iodent elites (E1 – E10), **Candidate donors** = 57 Iodent from the diversity panel

Identify and prioritize the Genetic Resources to use for Elite population enrichment

Increase of the overall progeny expected performance (donors ranked by introduction order)

Perspectives:
- Multi-trait approach: agronomic flaws in genetic resources
- Haplotype definition accounting for recombination frequencies
- Further simulations to validate the approach at short- and long-term in a breeding program
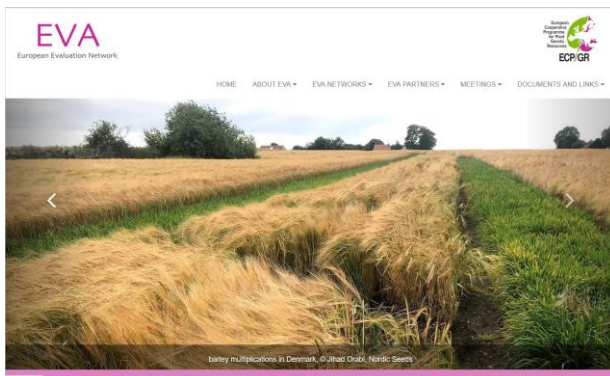
Conclusions

✓ ***Assessment of diversity and variance evolution in breeding is important to detect situations where they are at risk***

✓ Introduction of diversity is important to sustain genetic gain. ***Use of diversity donors needs to be adjusted given their gap in performance with elite materials.***

✓ Interest of phenotyping and genotyping Genetic Resources collections AND more recent materials (ex. Cooperative panels)

-> ***to detect favorable sources of diversity by training genomic prediction models***
-> to conduct association genetics and identify alleles of interest for traits already selected AND new traits (e.g. climate adaptation)

✓ Advances in genotyping technologies offer opportunities to detect under-used Genetic resources (e.g. Maize landraces)

✓ Great promises of DH technology to develop new inbred lines for exploiting landraces diversity (Hoelker et al., 2019)

Perspectives

✓ Application of previous approaches to adaptive traits and target environmental scenarios (e.g. combination with Millet et al., 2019, Nat Genet.)

✓ Practical implementation to address climate change and new objectives in agriculture,

-> need for pilot experimental programs and scale up, with an important potential role of multi-actor programs

https://www.ecpgr.cgiar.org/european-evaluation-network-eva

Acknowledgements

Thank you for your attention