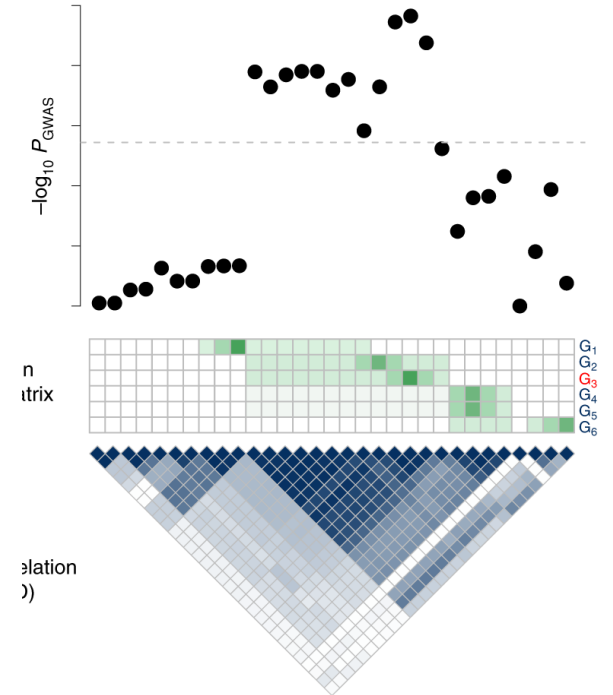
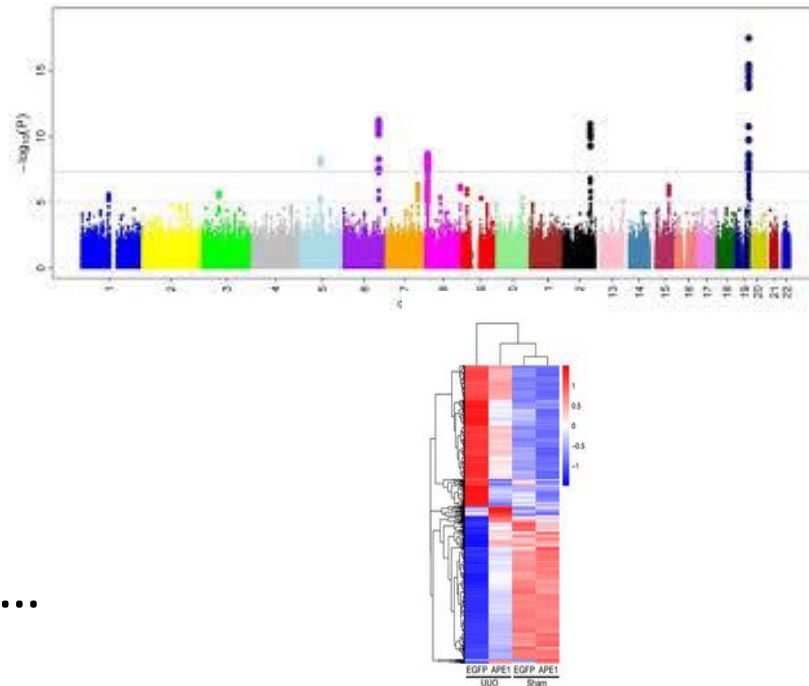


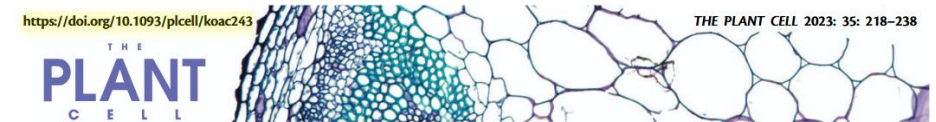
“Lessons” from the first two days



A single genotype, 3 ideotypes...



2 weeks => the whole summer
???



BREEDIT: a multiplex genome editing strategy to improve complex quantitative traits in maize



$$\Delta G = \frac{i r \sigma_g}{T}$$

Harnessing Global Diversity and Predictive Biology to Accelerate Sorghum Breeding

$$\Delta G = \frac{i \ r \ \sigma_g}{T}$$

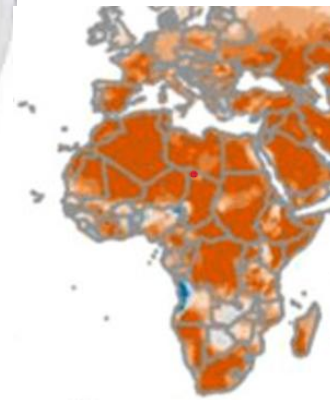
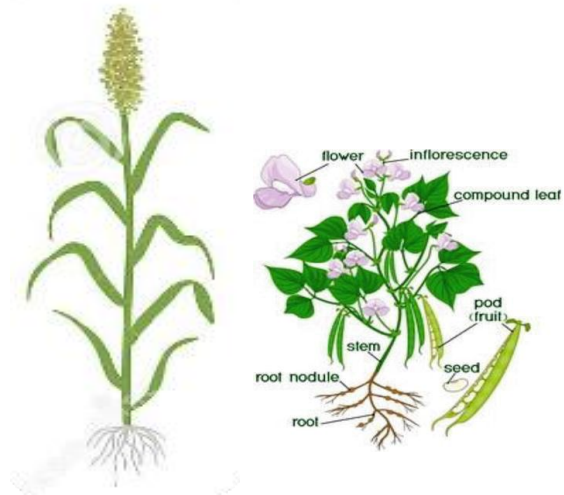
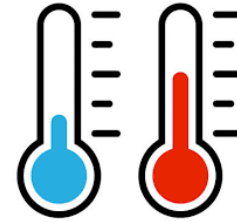
GxExM



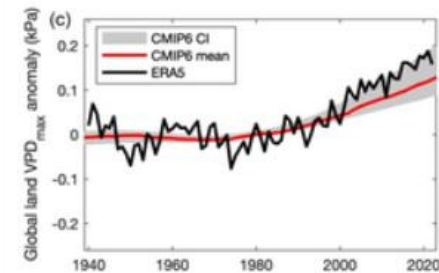
David Pot, Vincent Garin, Clément Bienvenu, Vincent Segura, Hugues De Verdal, Florian Larue, Gregory Beurier, Concetta Burgarella, Aude Gilabert, Angélique Berger, Korotimi Thera, Michel Vaksman, Niaba Teme, Jean-François Rami

The same challenges between EU and South partners

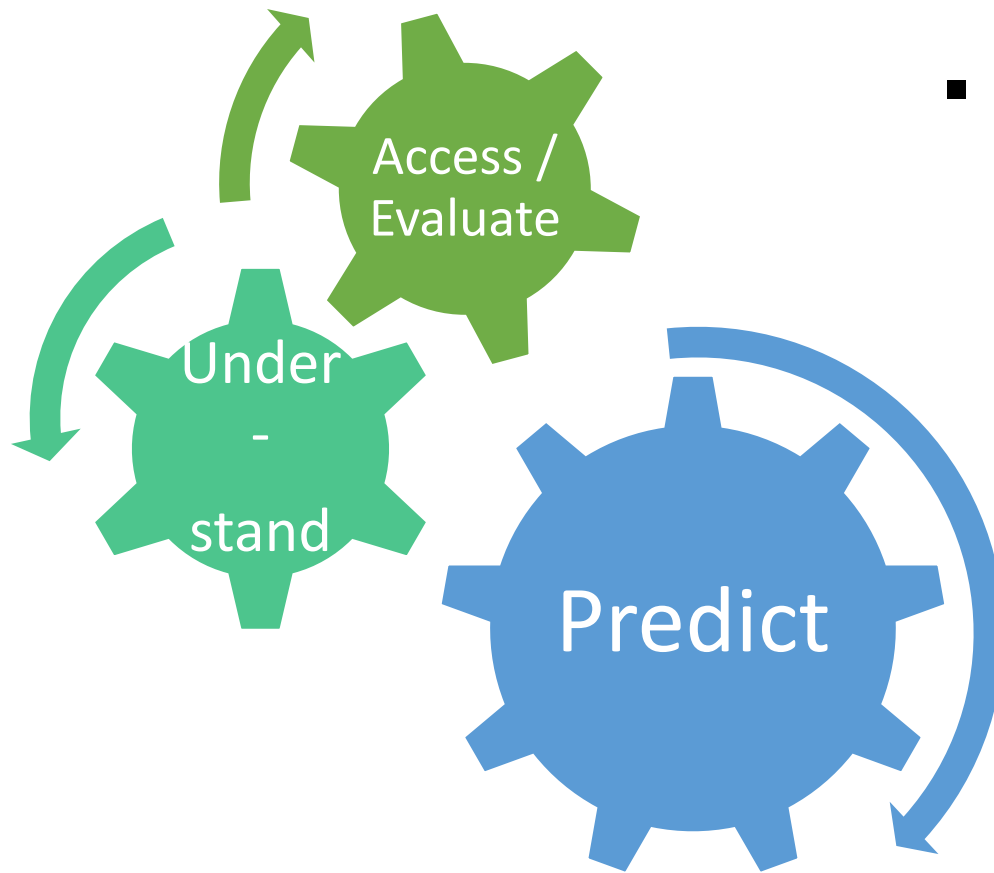
Similarity of challenges, but also an
« anticipation strategy »: Environmental
constraints, crop management
homologs....



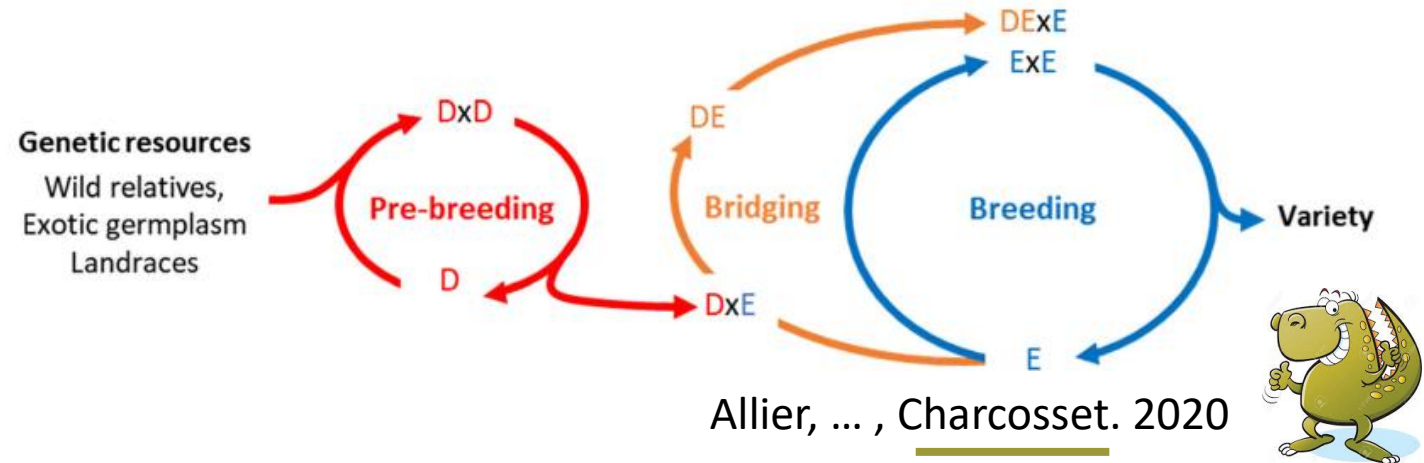
Vapor Pressure
Deficit



Three pillars (at least) to develop “adapted” varieties



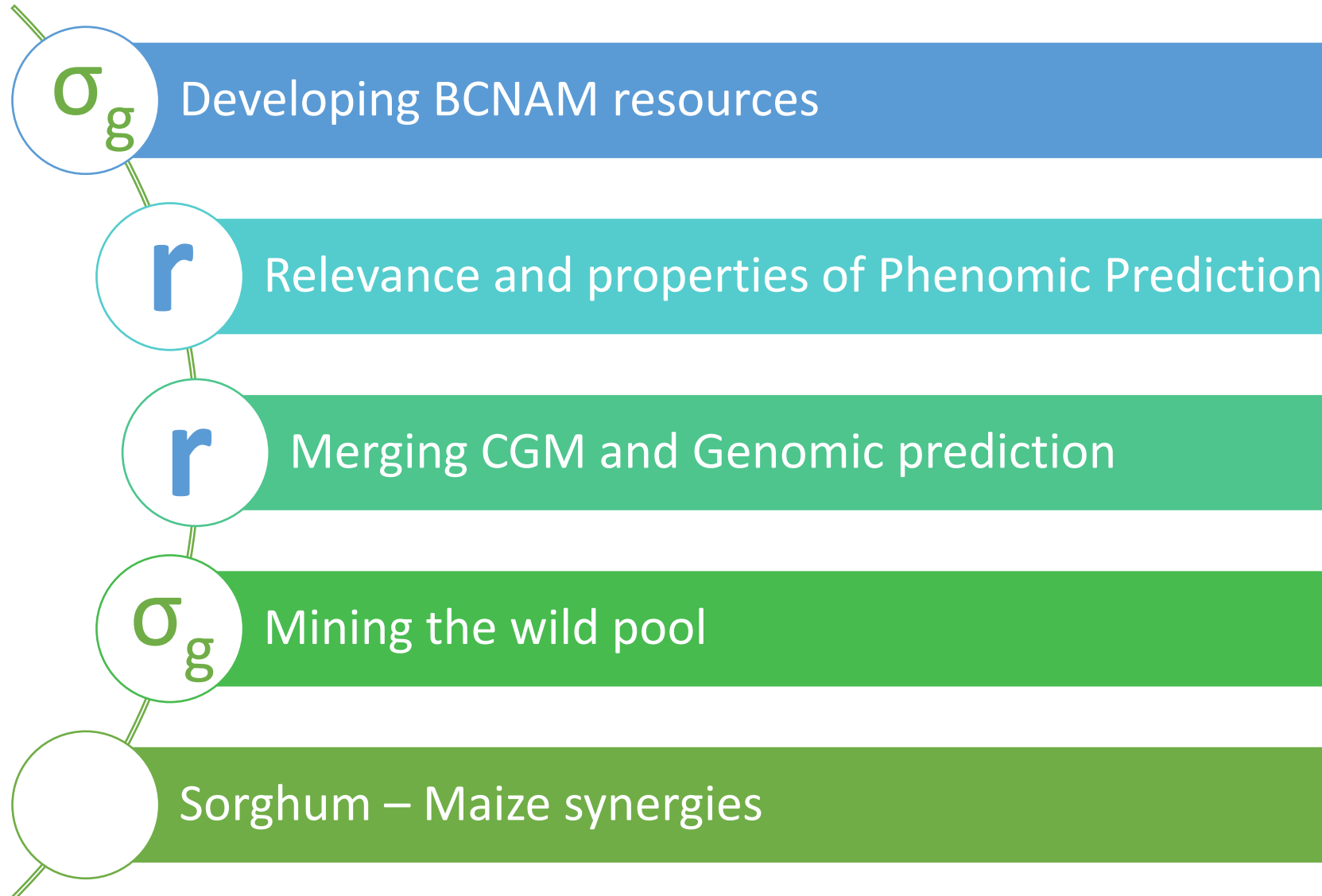
- Accessing the available diversity σ_g in relevant « genetic / environment / management » contexts still compulsory



$$\Delta G = \frac{i \quad r \quad \sigma_g}{T}$$

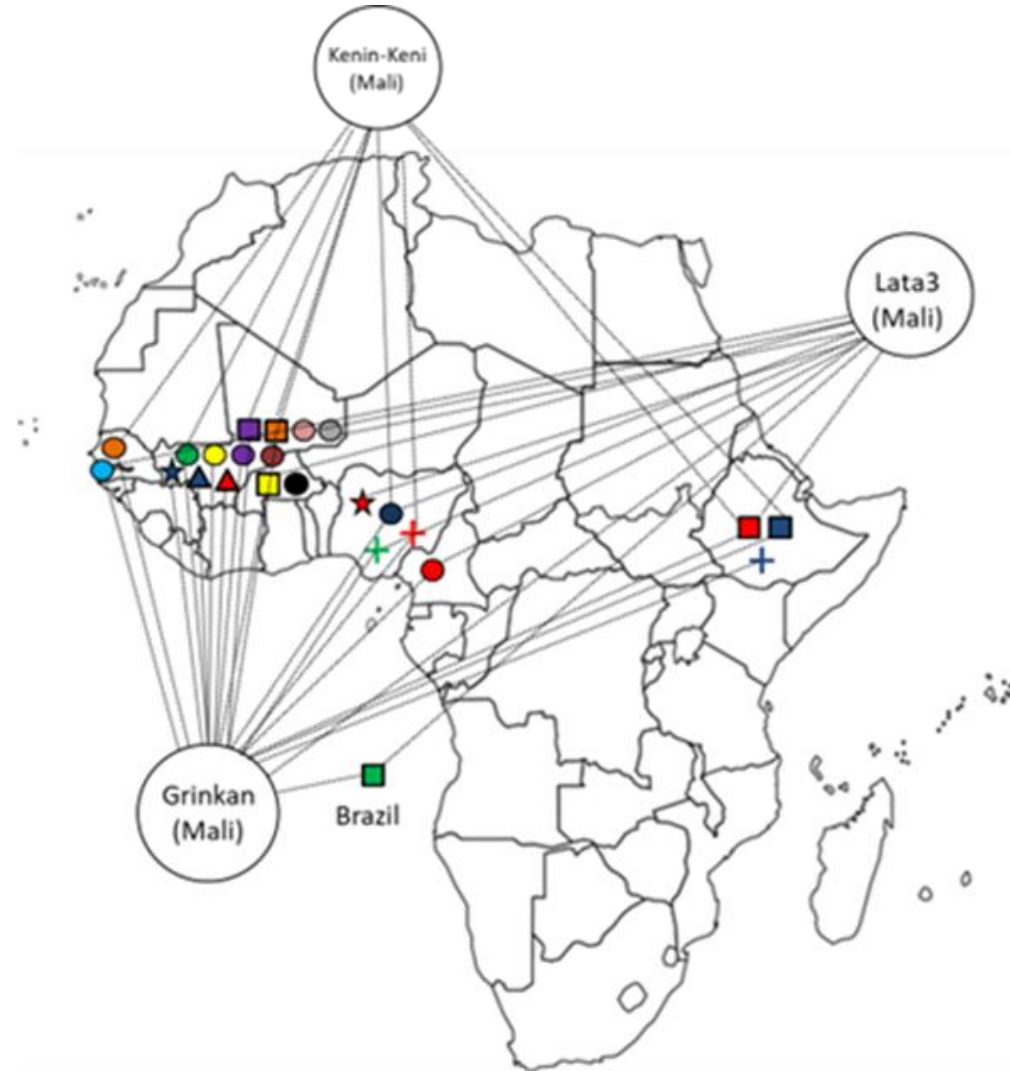
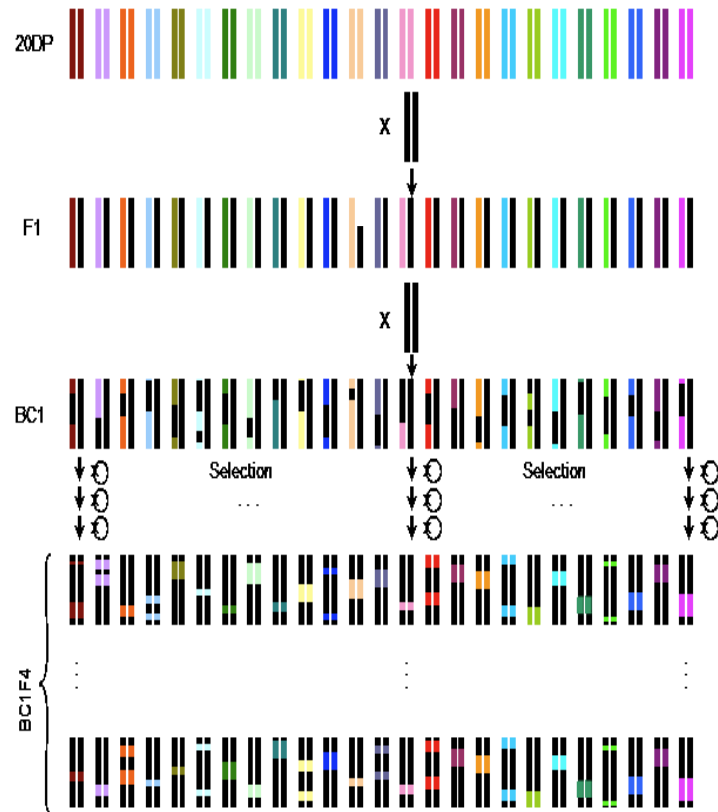
- « Understanding » remains a lever to optimize prediction accuracy r

Talk's outline



Developing a West and Central Africa BCNAM population

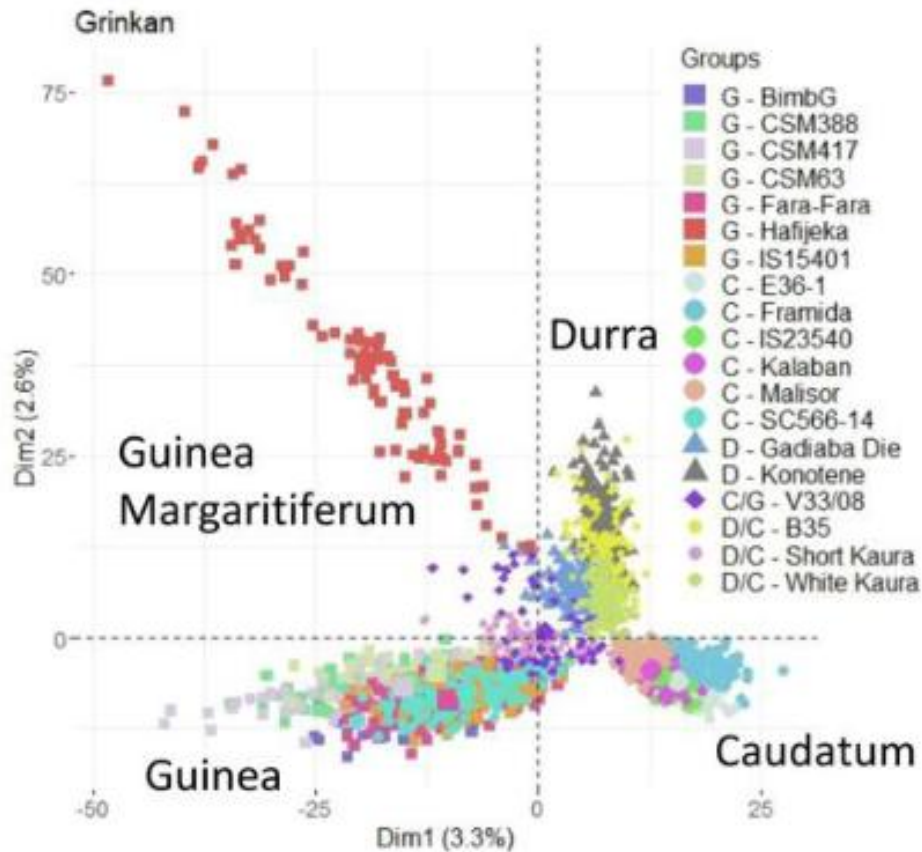
- BCNAM Jordan et al., 2011



- Fara-Fara (Nigeria)
 - IS15401 (Cameroun)
 - CSM417 (Mali)
 - CSM63 (Mali)
 - CSM388 (Mali)
 - BimbG (Guinea)
 - Hafijeka (Gambia)
 - Sangatigui (Mali)
 - DouaG (Mali)
 - N'Golofing (Mali)
 - Gnossiconi (Burkina Faso)
 - E-361 (Ethiopia)
 - IS23540 (Ethiopia)
 - SC566-14 (Brazil)
 - Framida (Burkina Faso)
 - Kalaban (Mali)
 - Malisor (Mali)
 - Konotene (Mali)
 - Gadiaba (Mali)
 - B35 (Ethiopia)
 - White Kaura (Nigeria)
 - Short Kaura (Nigeria)
 - V33/08 (Mali)
 - Sambalma (Nigeria)
- Guinea
- Caudatum
- Durra
- Gu/Caud Du/Caud

- 3901 BC1F3:5 families from 3 Recurrent and 24 donor parents

Developing a West and Central Africa BCNAM population

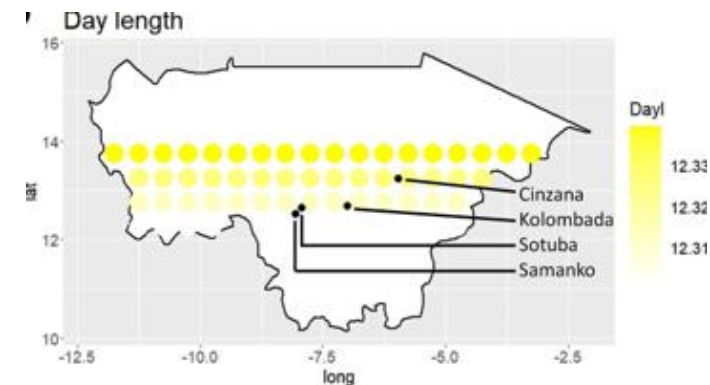
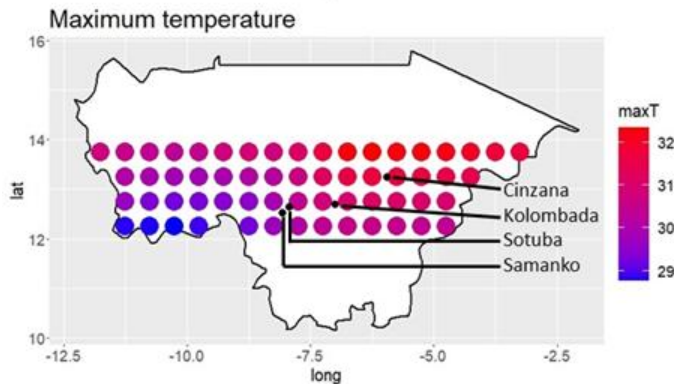
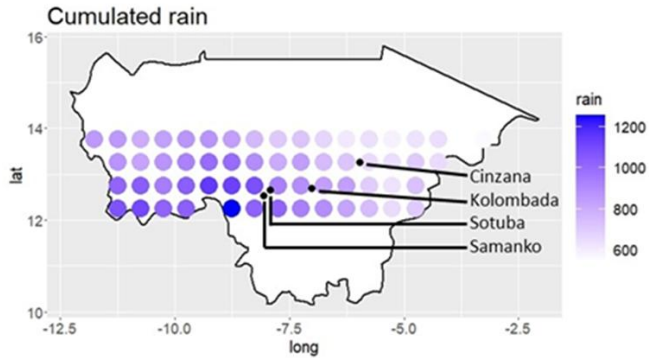


- WCA BCNAM: 69.9 to 86.8% of the global sorghum diversity
- US NAM: Bouchet et al: 38.12–57.5%
- Recurrent parents
 - WCA BCNAM : photoperiod sensitive
 - US NAM: photoperiod **I**nsensitive
- Complementarity of the WCA and US (BC)-NAM populations
 - Allele effects in different types of genetic backgrounds

Multi-environment phenotyping in Mali

■ Experimental conditions

- Four testing sites: Sotuba, Cinzana, Samanko, Kolombada
- Two sowing dates (Sot, Cin): End June, End July
- Low and High P (Sam)

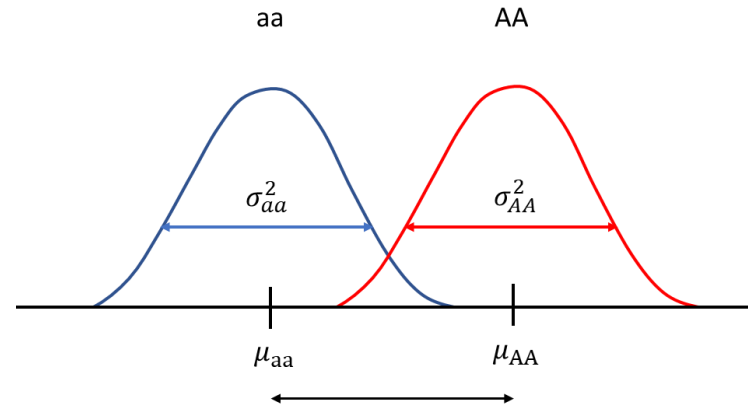
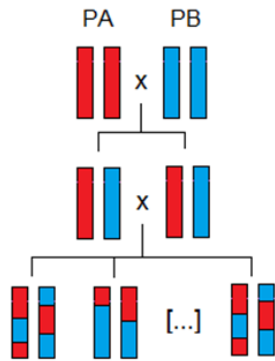


■ Trait:

- Flag leaf emergence
- Plant height
 - Number of internodes
 - Average internode length
 - Peduncle length
- Panicle length
- Yield

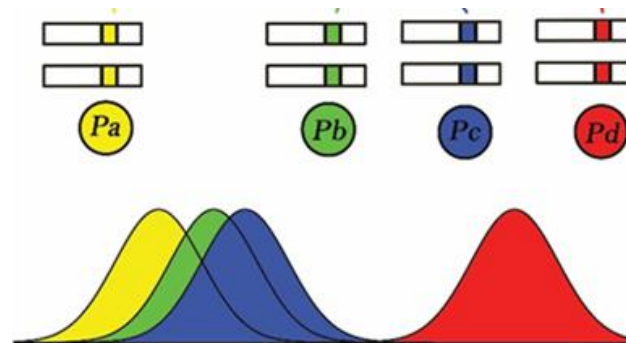
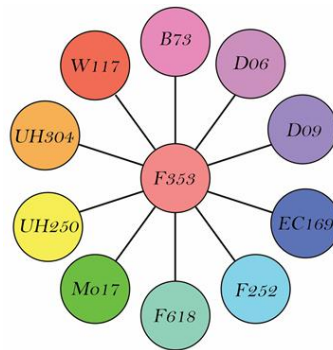
QTL detection in Multi-parent Populations

Bi-parental
population



$$y_i = \mu + x_i\beta + \epsilon_i$$

Multi-parent
Population (MPP)



$$y_{ic} = \mu + C_c + x_{ip}\beta_p + \epsilon_i$$

Vincent Garin,
Garin et al., 2024



QTL: Genetic background and Environment dependencies

- 5 « Recurrent Parents x Env » analyzed
- 100 QTL => 64 Unique ones

FLAG, PH, IN length : 10-48 %

PED, Node number, 100 Grain Weight : 8-30 %

Yield, Panicle length : 4-14 %



	FLAG
Grinkan 2012	6 (48.9)
GR 2013	4 (32.1)
KK 2012	6 (53.4)
KK 2013	2 (35.5)
LT	4 (50.3)

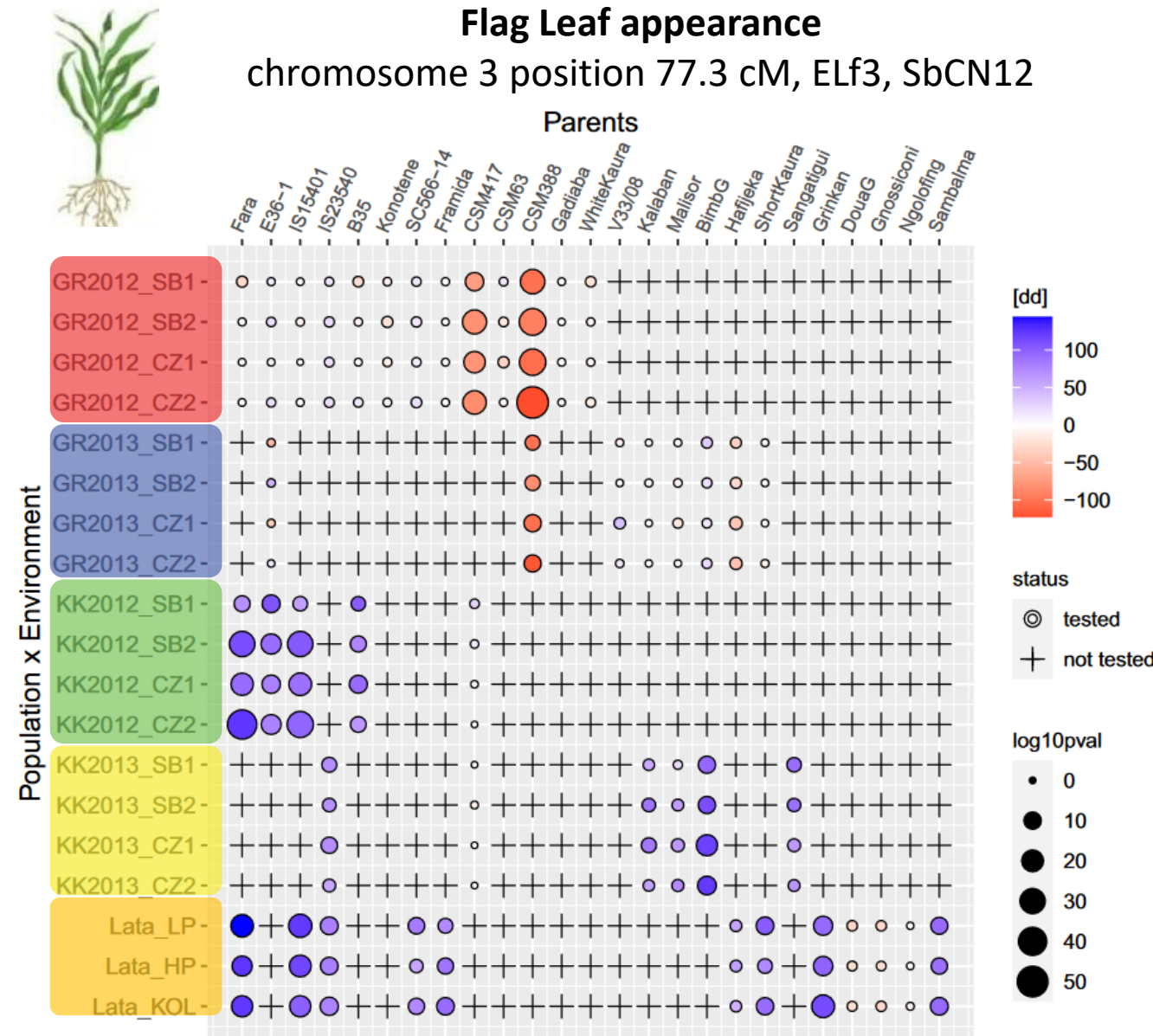
22 QTL

Q_FL_3_78 => Elf3, SbCN12

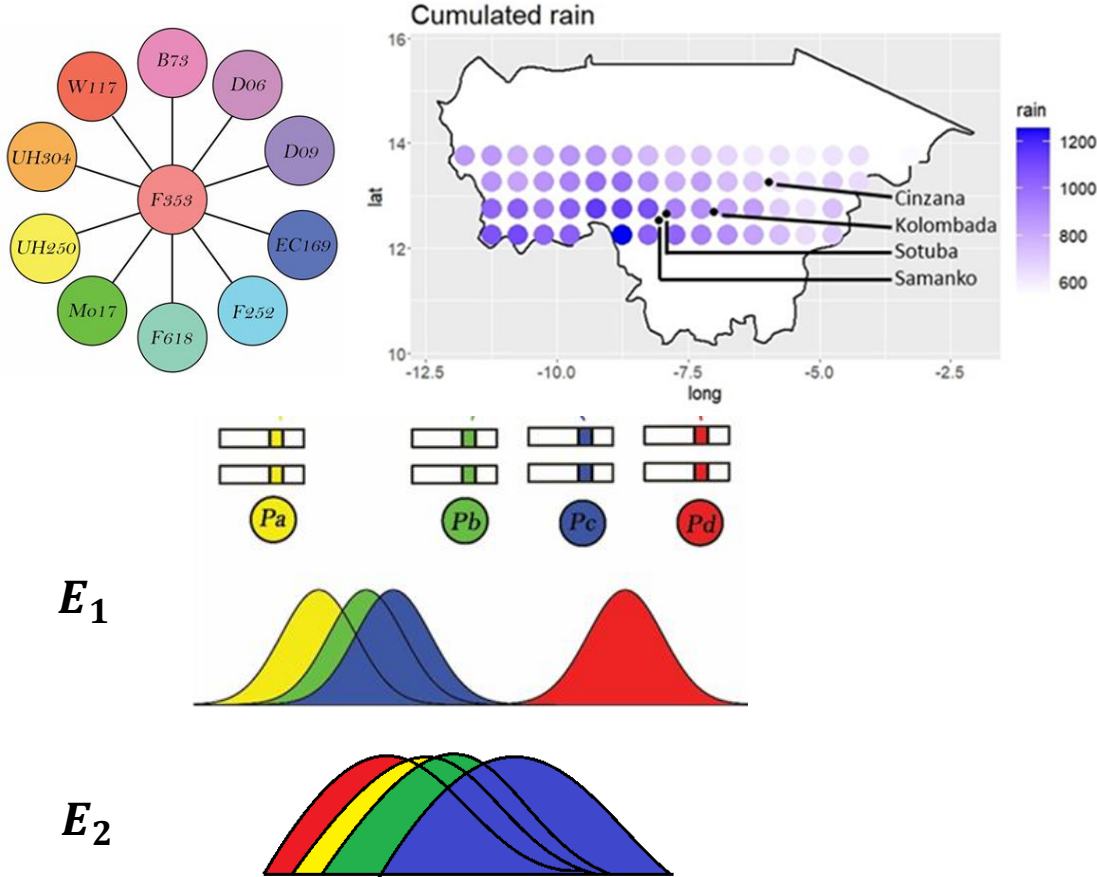
Q_FL_6_3 => Ma6

Q_FL_6_38 => Ma1

Q_FL_9_105 => SbFL9.1



Diving in the G x E



$$y_{icj} = \mu + e_j + c_{cj} + x_{ia} * \beta_{aj} + \underline{ge}_{icj} + \underline{\epsilon}_{icj}$$

■ Globally

- 100 QTL => 948 parental alleles with potential effects
- 51% of the parental alleles presented significant effects (/RP)
- 26% interacted with the environment (238 alleles corresponding to 87 QTL)

■ Flag leaf

- 22 QTL
- 200 parental alleles, 100 with significant effects
- 22 % interacting with the environment (GXE)



From G x E to G x Environmental Covariates

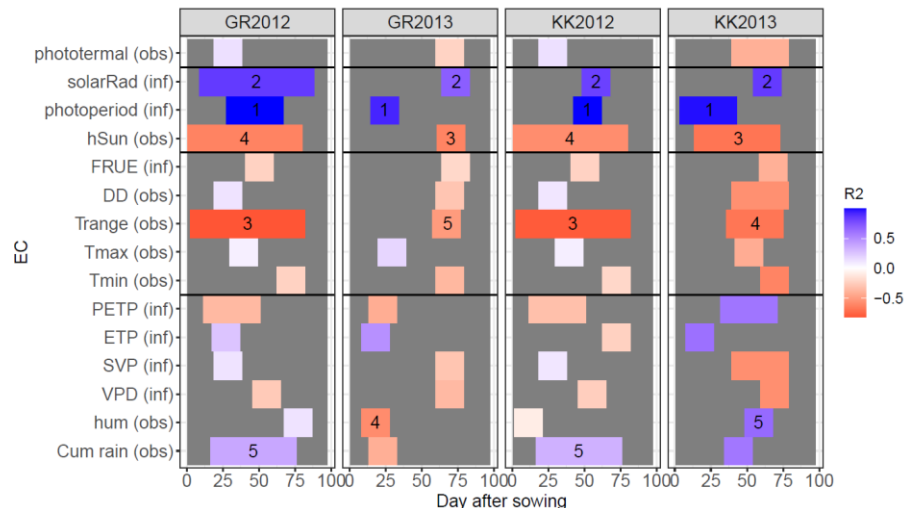


- Local envirotyping
- NASA Power

$$\underline{y_{icj}} = env_j + cross_{cj} + \underbrace{x_{ip} * \beta_{pj}}_{\text{QTL term}} + \underline{GE_{icj}} + \underline{e_{icj}}(2)$$

QTL term

$$x_{ip} * (\beta_p + EC_e * S_p + l_{pe})$$



Globally

- 16% QTL alleles interacted with at least 1 EC
 - 128 alleles corresponding to 61 QTL

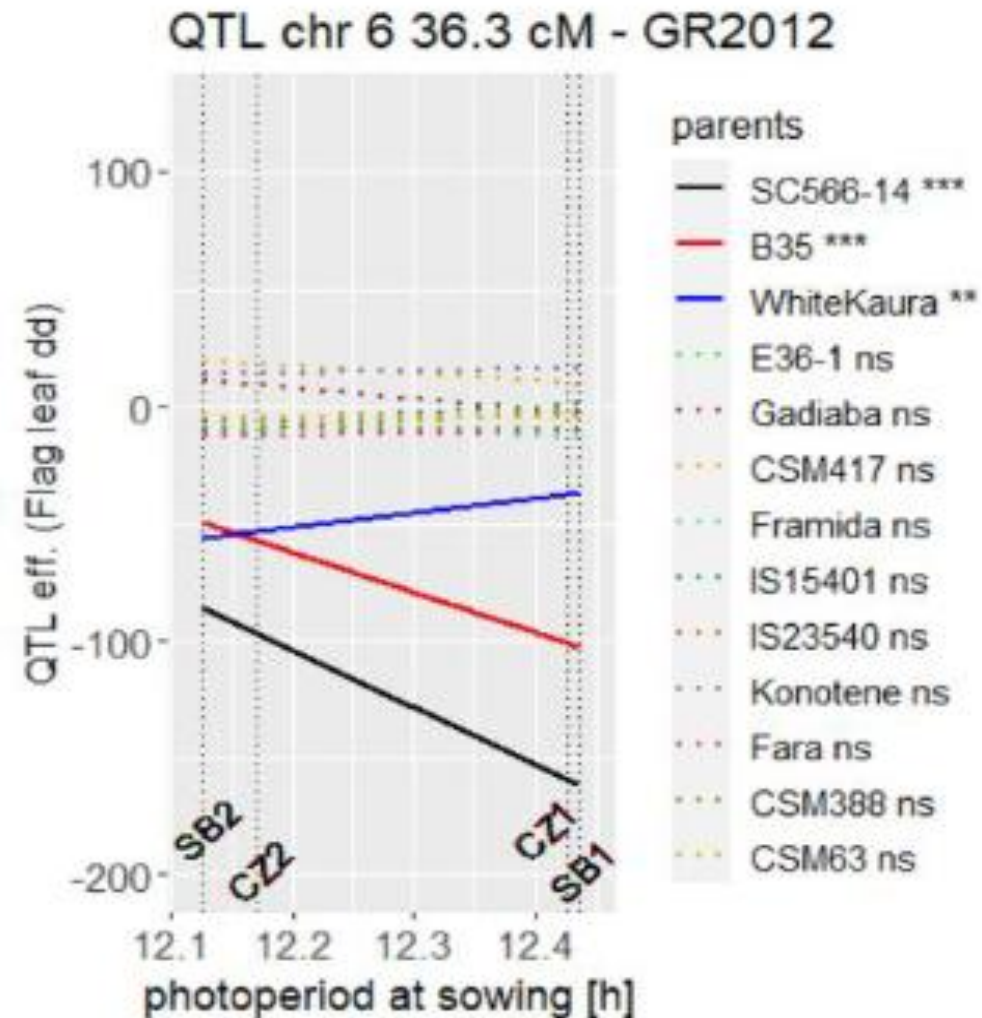
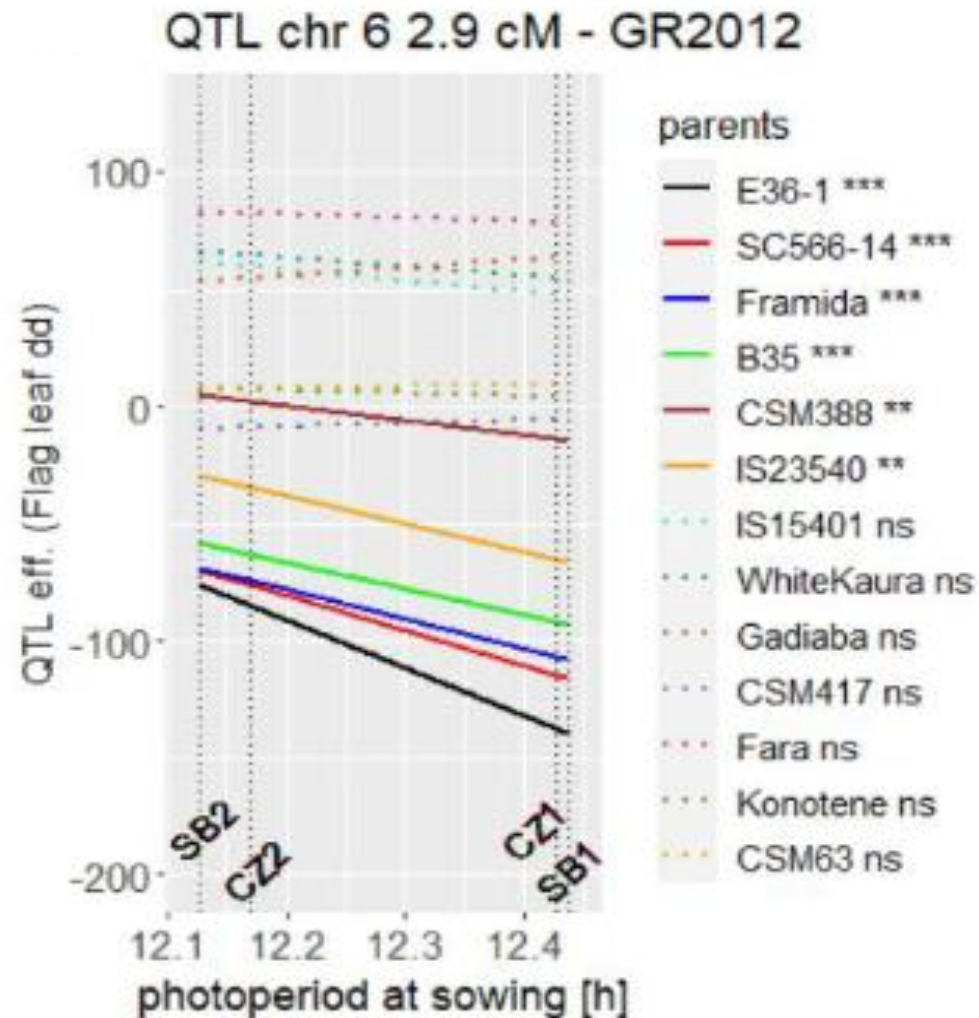
Flag leaf

- 30 parental alleles (15 %) interact with environmental covariates
 - Rain : 21
 - T range : 23
 - Hsun : 23
 - Photoperiod : 24
 - Solar Rad : 23
 - Air humidity : 5
- Total : 119 interactions involving 30 alleles**

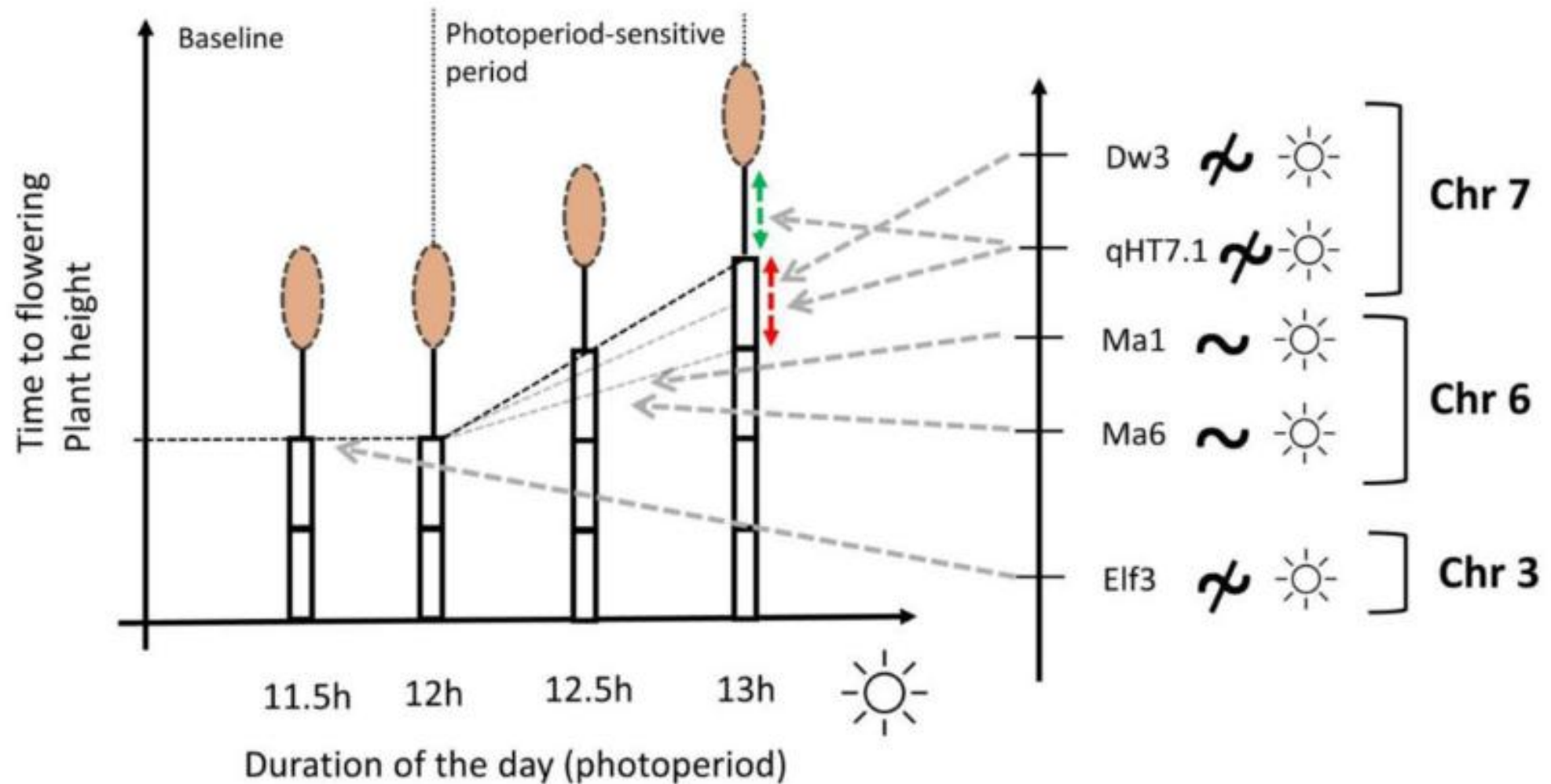
Flag leaf emergence: Better understanding G x EC

Ma6

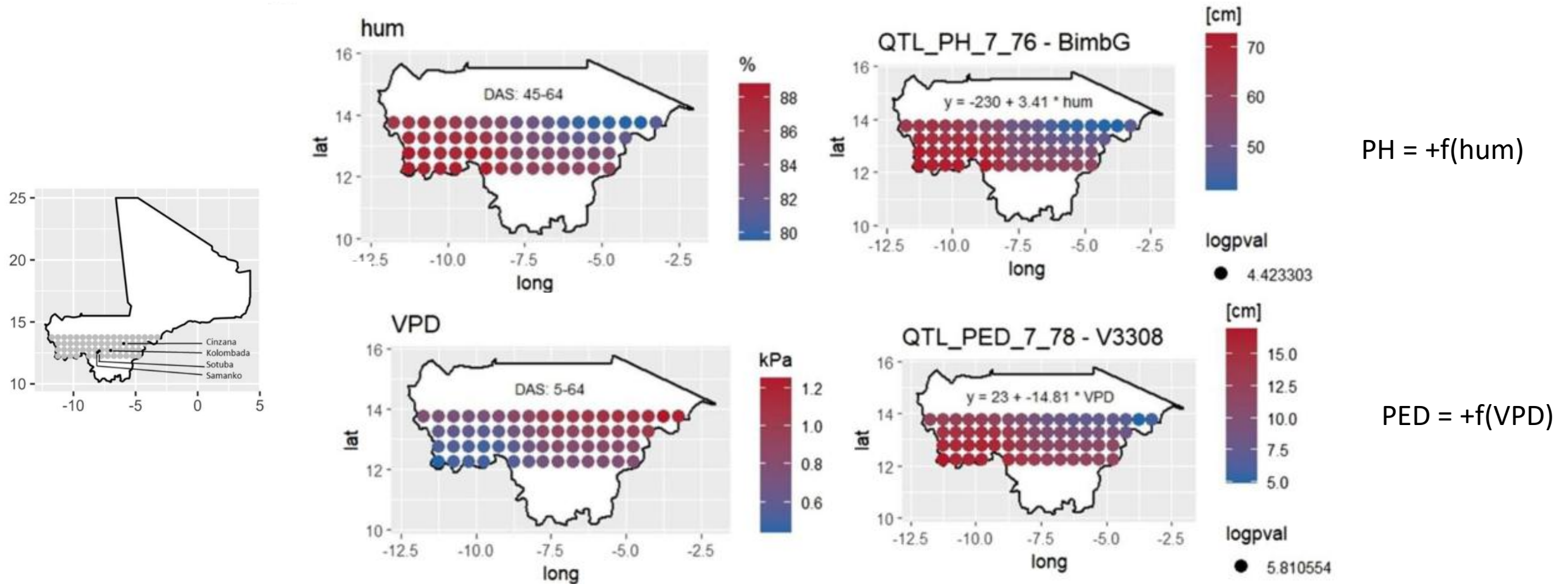
Ma1



Flag leaf emergence: Photoperiod dependent and independent



Extrapolating QTL effect beyond the tested environments

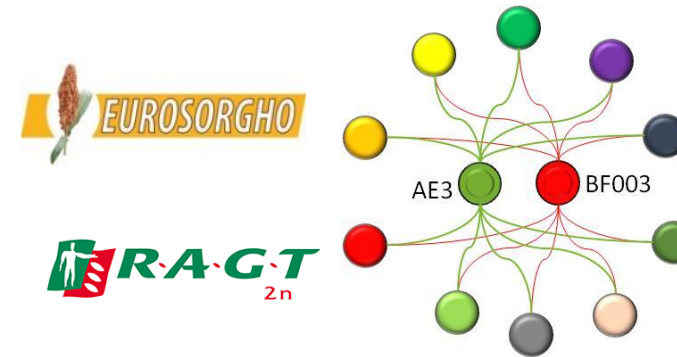


- Gives an idea of the interests / effects of the QTL alleles in untested environments
- **BUT it has to be validated for larger uses**

BCNAM X Env X EC : Take home message

- BCNAM populations allow to test allele effects in a common genetic background
- Relevant for genetic and breeding goals
- Allowed a better understanding of the genetic determinism of phenological and architectural traits

- EU Biomass BCNAM



2 RP
10 DP
2000 BC1F4:6

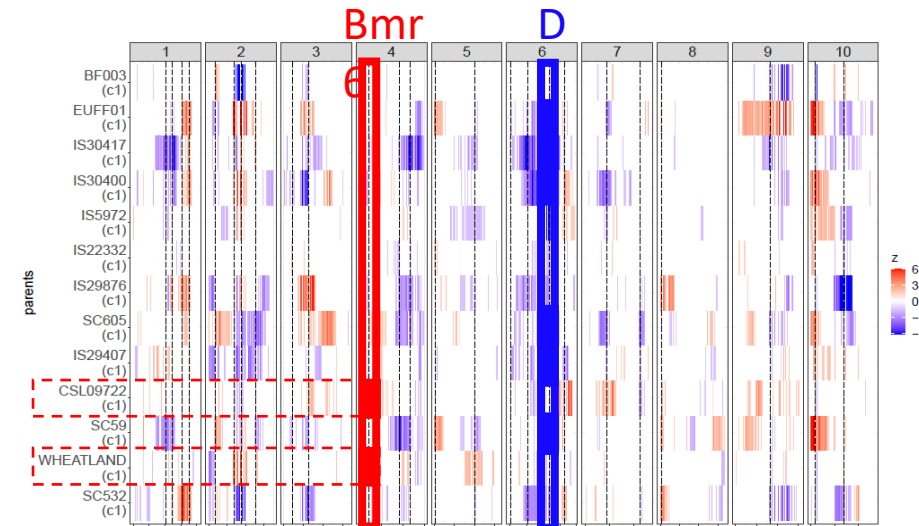
- New traits being analyzed in these populations: root angle, root soil agregation in Western Africa (WCA BCNAM) and EU (EU Biommass BCNAM)



Laurent Laplaze IRD France



Isabelle Basile-Doelsch, INRAE, France



Talk's outline

σ_g

Developing BCNAM resources

r

Relevance and properties of Phenomic Prediction

r

Merging CGM and Genomic prediction

σ_g

Mining the wild pool

Sorghum – Maize synergies ?

Prediction of Genetic estimated Breeding Values

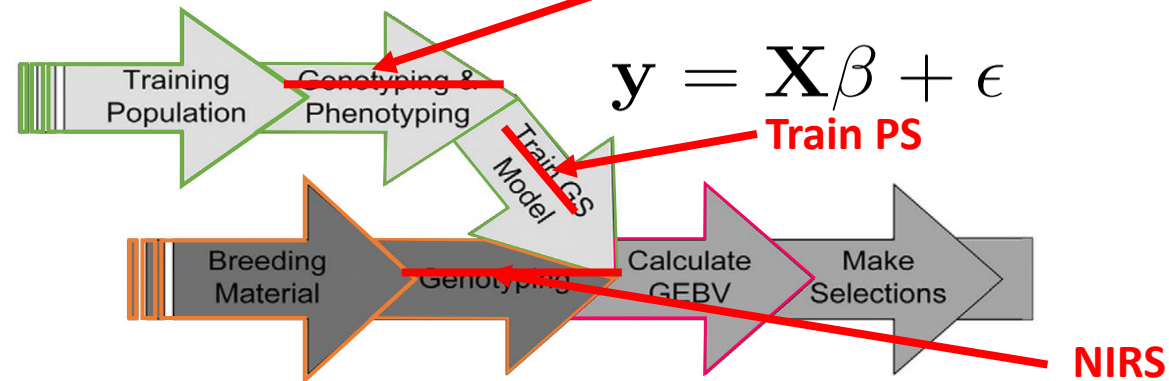


Genotype	T1	T2	...	M n
1	10	1.7		
2	12	2.1		
3	15	1.5		
4	14	3.2		
5	16	4.1		
...				
n	20	0.8		

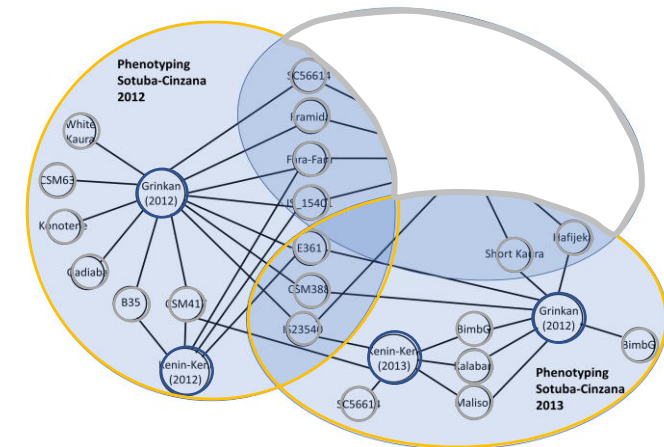


Rincent et al., 2018

Genotype	M 1	M 2	M 3	M 4	M 5	...	M n
1	A	C	T	C	A		
2	T	C	G	C	G		
3	A	C	G	G	G		
4	A	C	G	G	A		
5	T	A	G	G	A		
...							
n	A	A	T	C	A		



- Subset of the WCA BCNAM
 - 2 recurrent parents
 - 22 donors
 - 2498 BC1F3:5 families



Accuracy of prediction = correlation between **predicted values** and **observed / measured ones**

Phenomic Prediction: 3 main questions

- Is phenomic Prediction useful in sorghum: in a context of « line » evaluation, in a multiparental population
- What are the effects of the training set size and population structure on phenomic prediction accuracy ?
- Do we really take advantage of the whole set of information available in the NIRS spectrum ?

Experimental set-up

GR =
1403 genotypes
13 crosses

KK =
418 genotypes
5 crosses

2012	Sotuba	Cinzana
Sowing date 1	ENV 1	ENV 2
Sowing date 2	ENV 3	ENV 4

GR =
528 genotypes
8 crosses

KK =
255 genotypes
6 crosses

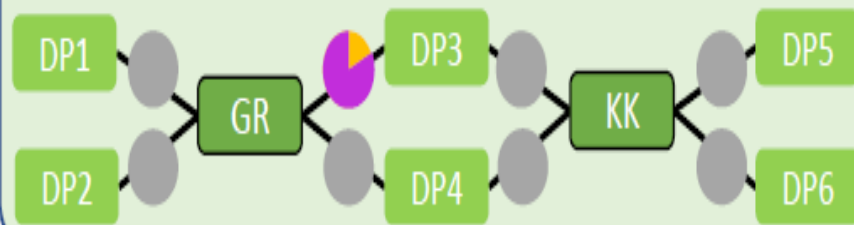
2013	Sotuba	Cinzana
Sowing date 1	ENV 1	ENV 2
Sowing date 2	ENV 3	ENV 4

■ NIRS acquisition environment

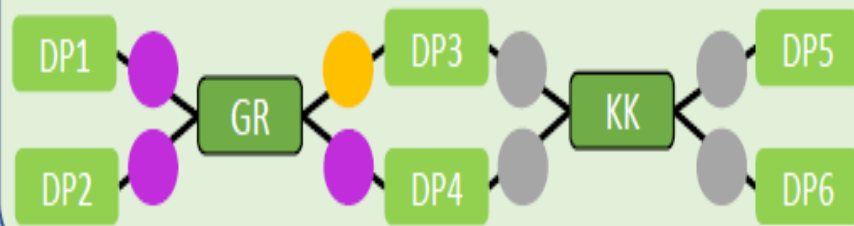
■ Breeding value estimation environment

Population structure effect:
decreasing relatedness bw Cva and Cvc

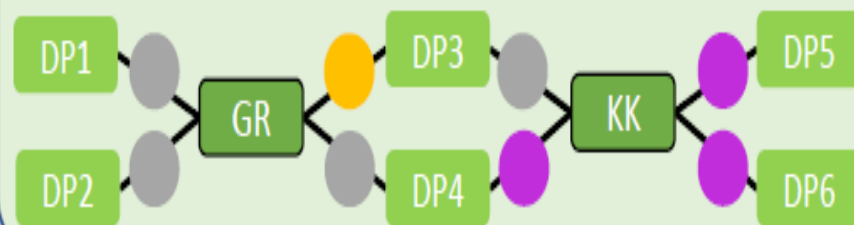
CVa: 5-fold cross validation per cross



CVb: 1 cross predicted by its recurrent half sib



CVc: 1 cross predicted by crosses that share no parents

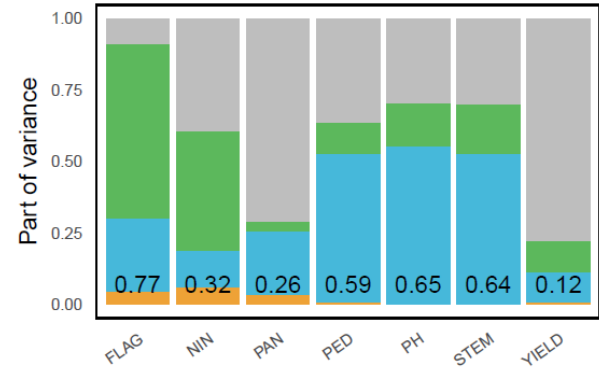


● Training set
● Validation set
● Not used

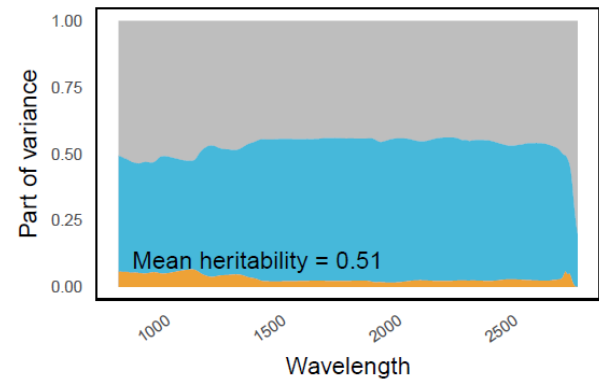
Training size effect

Comparing Phenomic and Genomic selection

2012

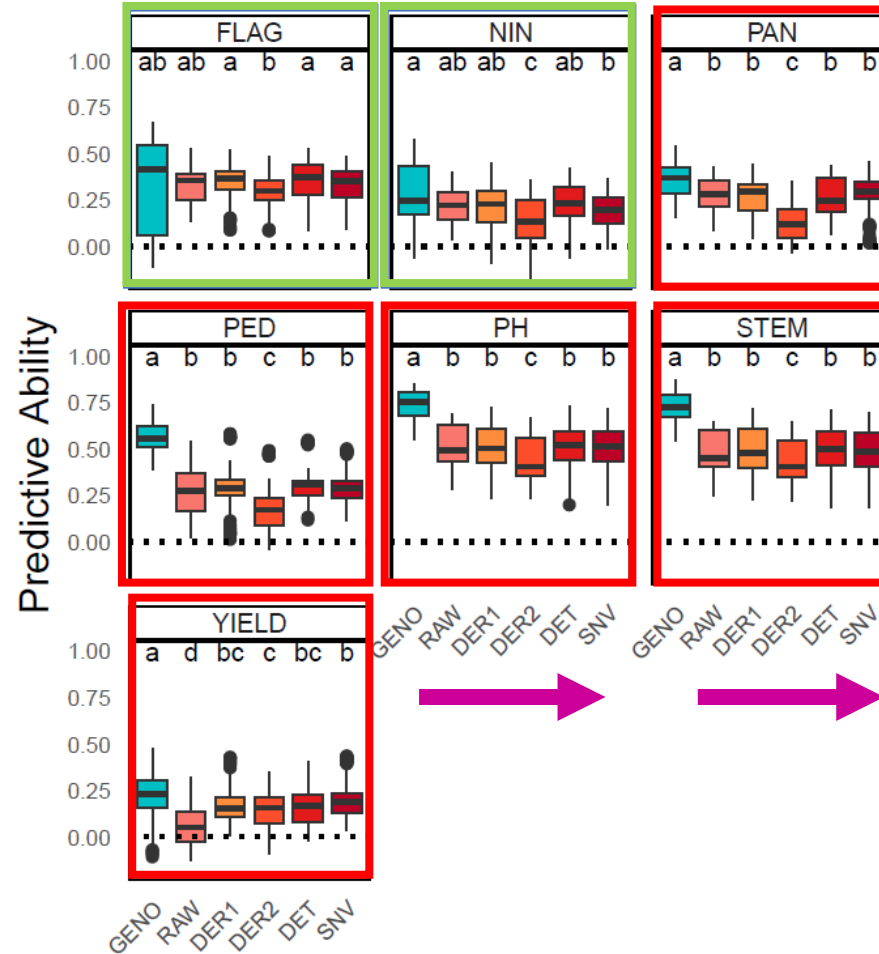


2012

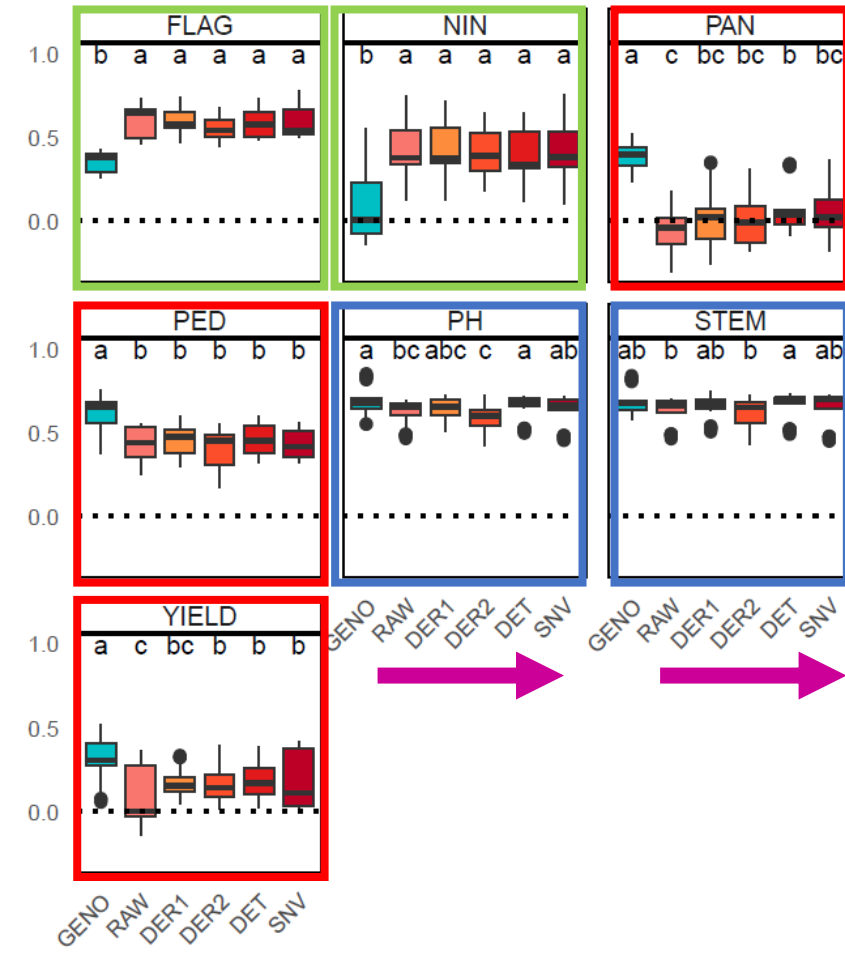


Legend : ■ RESIDUAL ■ ENV ■ GENO ■ CROSS

2012



2013



2012 : PP prediction accuracy **lower** of **equivalent** to GS

2013 : For some traits PP prediction accuracy **higher** than GS

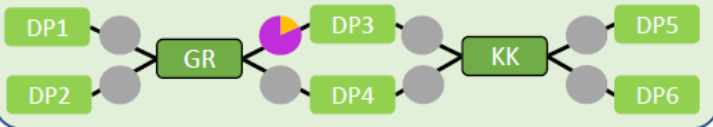
2013 : Almost no effect of the **spectrum pre-treatment methods** / prediction

PP works... Slightly lower than GP
Trait dependent

Effect of population structure

Population structure effect:
decreasing relatedness bw Cva and Cvc

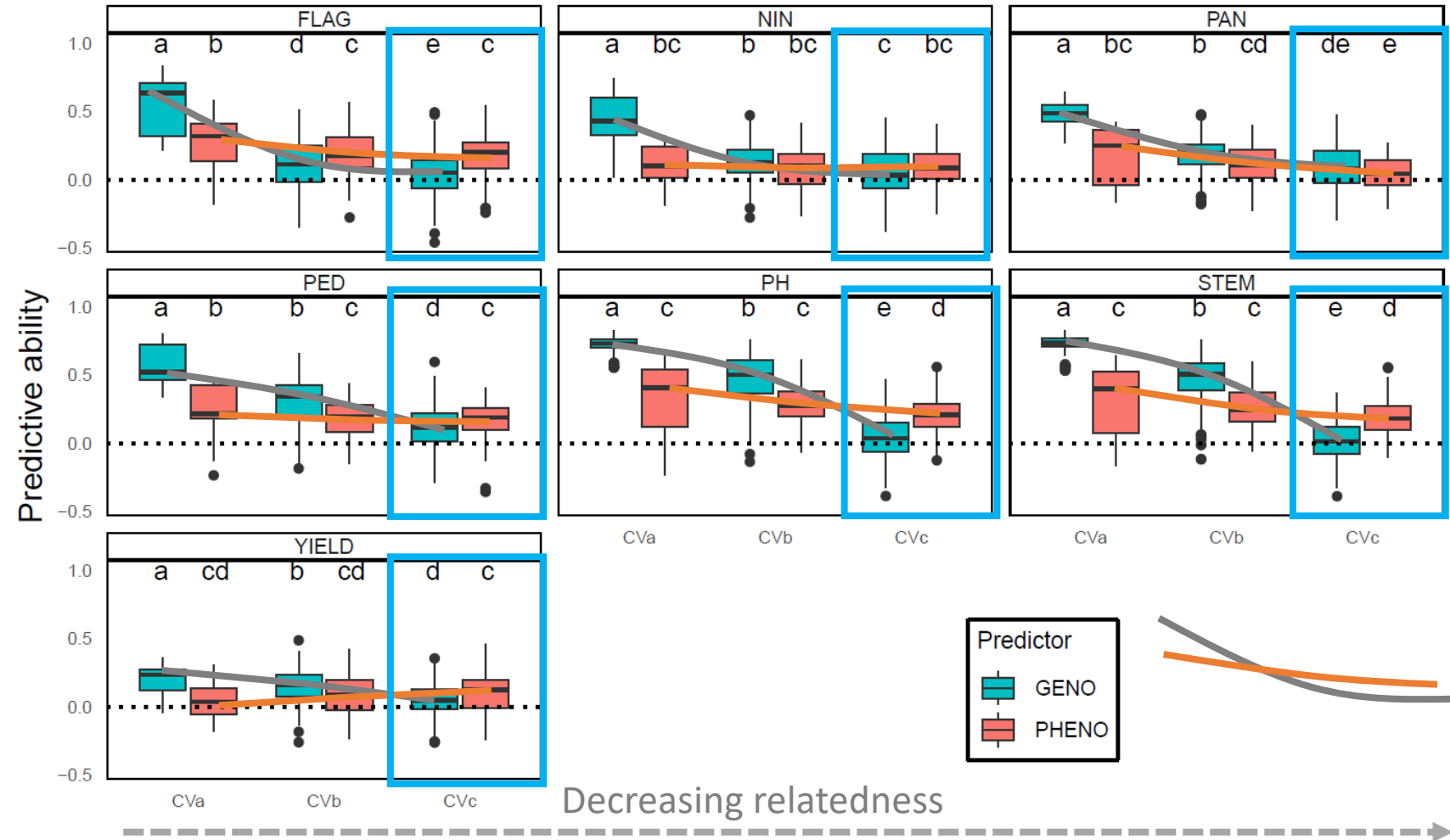
CVa: 5-fold cross validation per cross



CVb: 1 cross predicted by its recurrent half sib



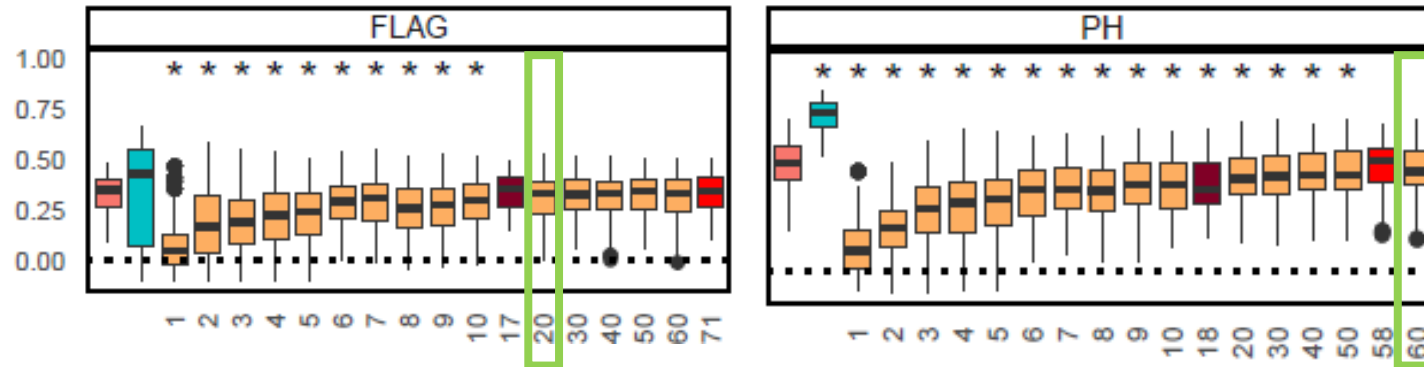
CVc: 1 cross predicted by crosses that share no parents



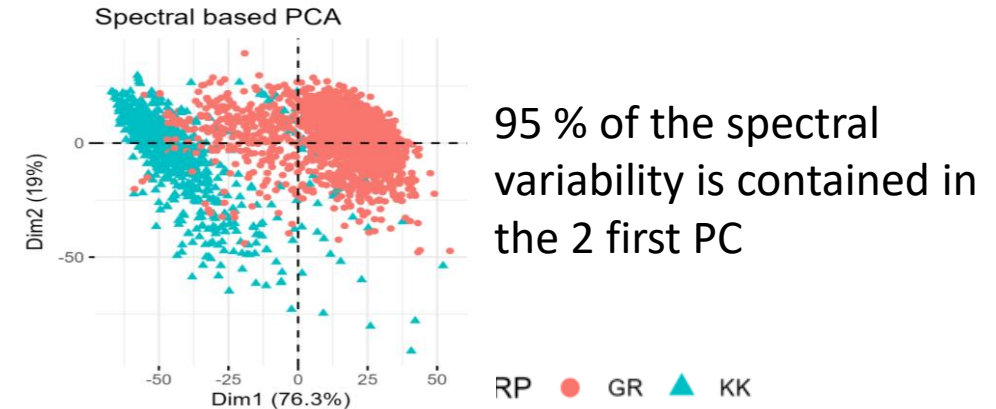
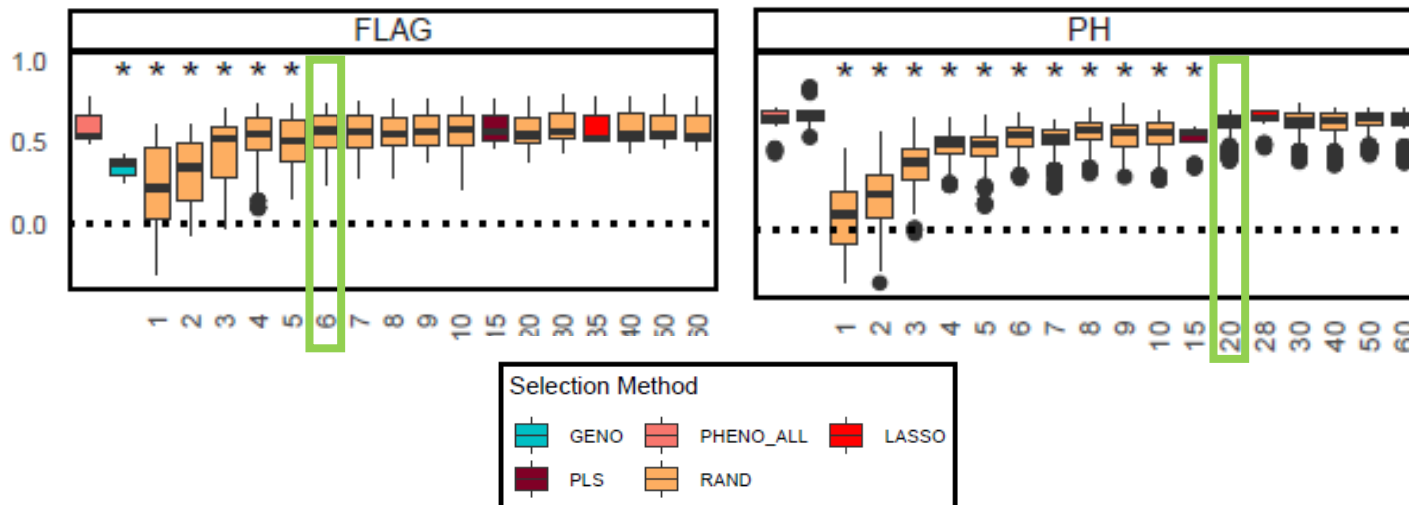
- PP is less affected than GP by population structure (in accordance with Laurençon et al...)
- PP had higher PA than GP for almost all traits in the scenario with the least relatedness between training and validation sets

Extremely small numbers of wavelengths allows to reach a maximal predictive ability

2012



2013



- Costs and benefits of using only a small set of wavelengths ?
- Current genetic approaches are likely not efficient enough to take advantage of NIRS information (based on individual wavelengths)
- New chemometrics methods recognizing highly correlated structure of spectral data are being explored aiming to better split G, E, and GXE components

Phenomic selection: Take home message

- Phenomic Prediction « works », PA slightly lower than GP but it is trait dependent...
- **PP: Higher transferability to distant population than GP**
 - Similar to results obtained with endophenotypes (RNA, Proteins, Metabolites)
 - Or models taking advantage of functional annotations (end of talk)
 - **Cases where only causative information matters**, LD is not yet a way to take advantage of distant causative genes
- Still space for large improvement through optimal wavelengths exploitation => Connection with chemometrics on-going
- Currently looking for large GXE sets of NIRS information !!!!



Clément Bienvenu
PhD: 2023- 2026

Paper under review

Talk's outline

σ_g

Developing BCNAM resources

r

Relevance and properties of Phenomic Prediction

r

GxExM: Merging CGM and Genomic prediction

σ_g

Mining the wild pool

Sorghum – Maize synergies ?

GxExM: Optimizing prediction accuracy

■ Increasing prediction accuracy

- Taking into account environmental information
- Taking advantage of available knowledge
- Improving the prediction method
- ...
- Combinations of all of these

■ Connecting CGM with Genetic and prediction analysis



François Tardieu's « school of thinking »

■ Crop Growth Models (CGM) and Functional Structural Plant Models (FSPM)

- EXM integrators
- Crop Growth Model: Phenology, Yield...
- FSPM: take advantage of 3D information

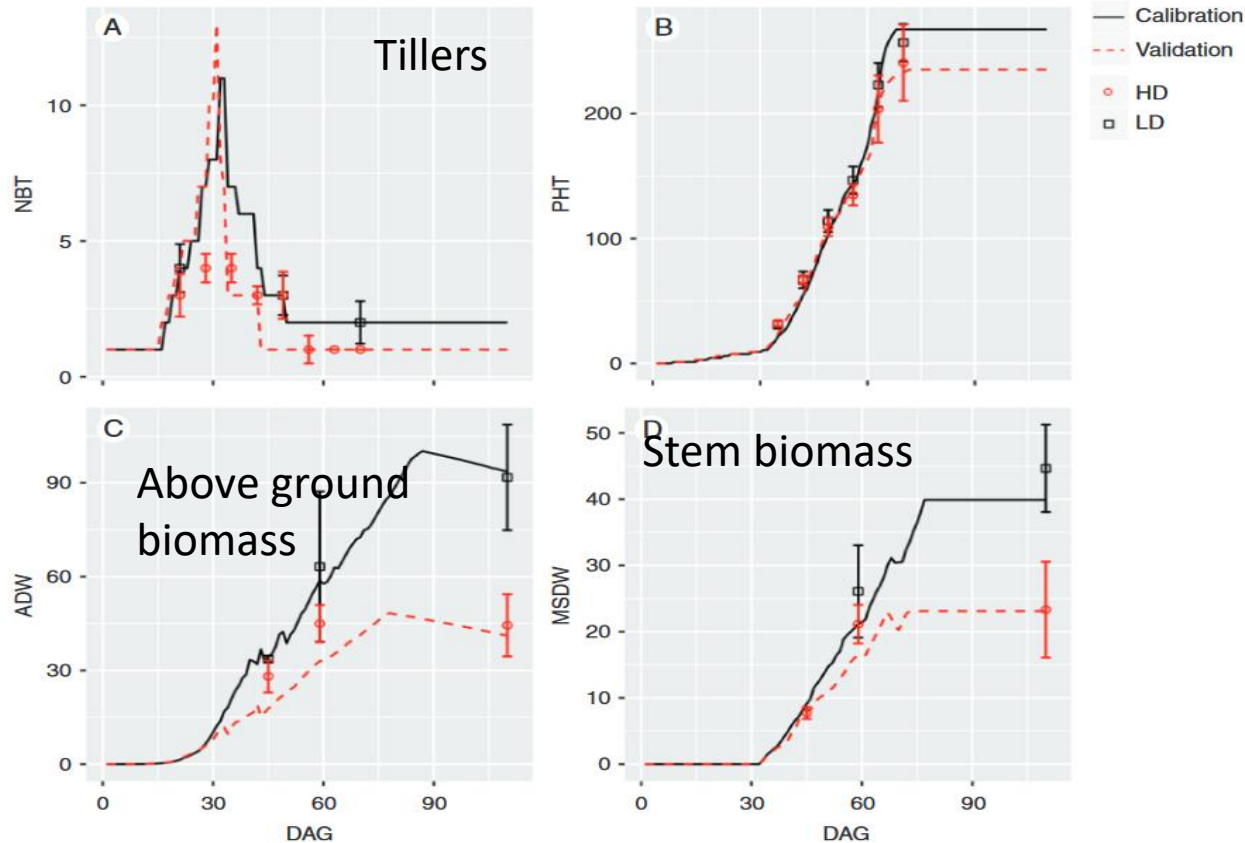
$$Y = f \left[\begin{array}{c} \text{Sun and Rain} \\ \text{Watering} \\ \text{Fertilizer} \end{array} \right], \text{ model parameters}$$

$f[]$: linear or non linear function

Model parameters : some of them depends on genetics

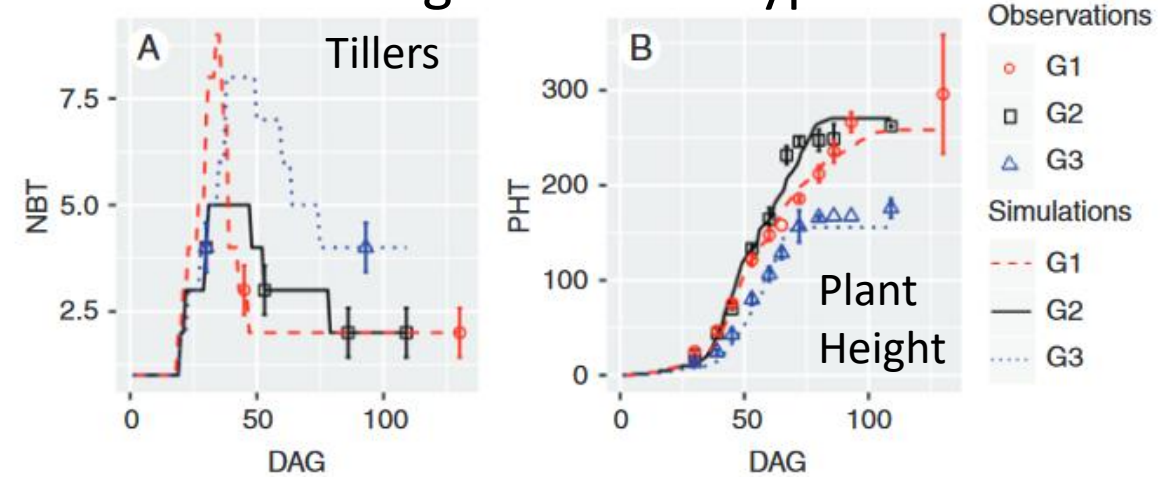
Sorghum EcoMeristem: outputs

■ Dealing with management, low vs high density

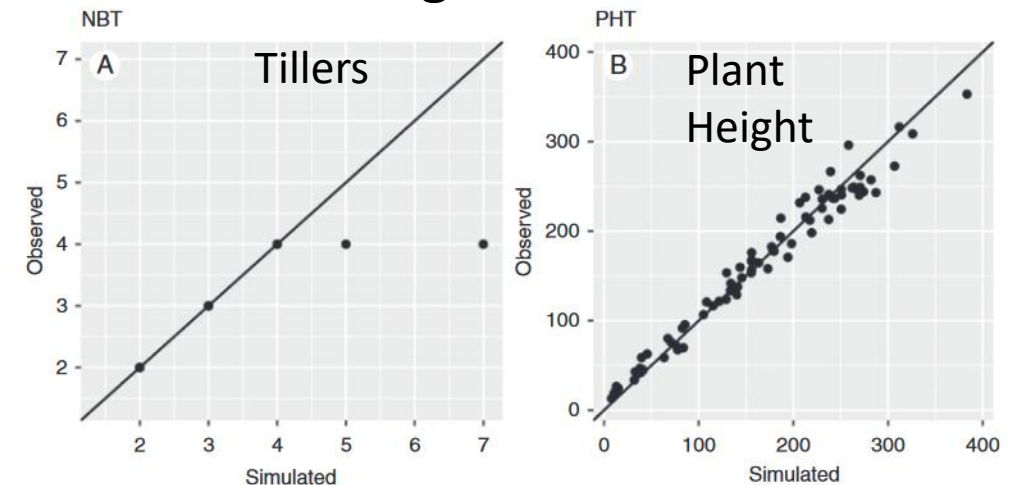


Florian Larue (PhD)
Larue et al., 2019

■ Dealing with Genotypes

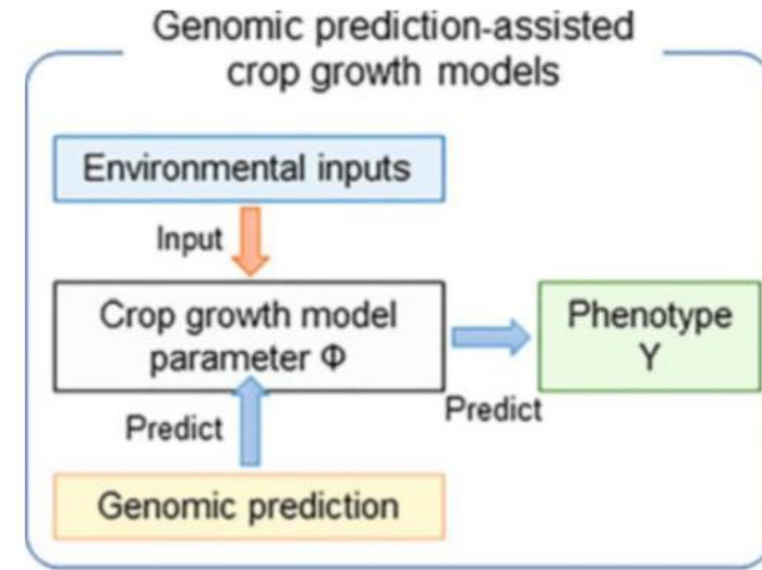
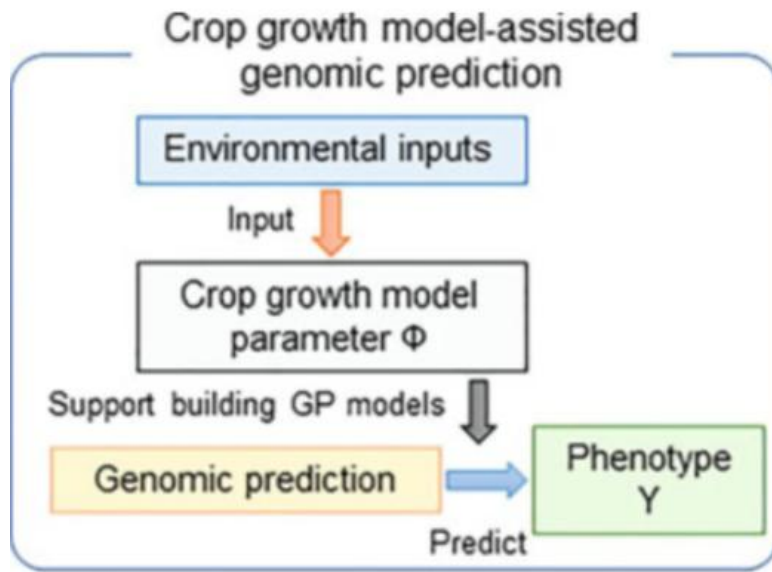


Dealing with Env Effects



Training 2014, Test 2015 (8 genotypes)

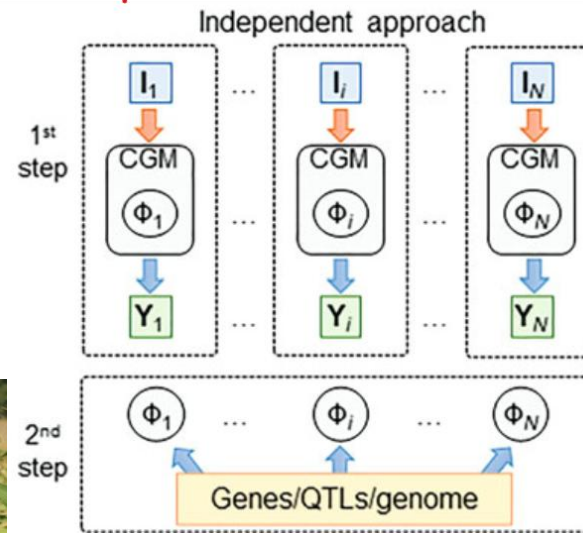
Coupling CGM / FSPM with GP (/QTL)



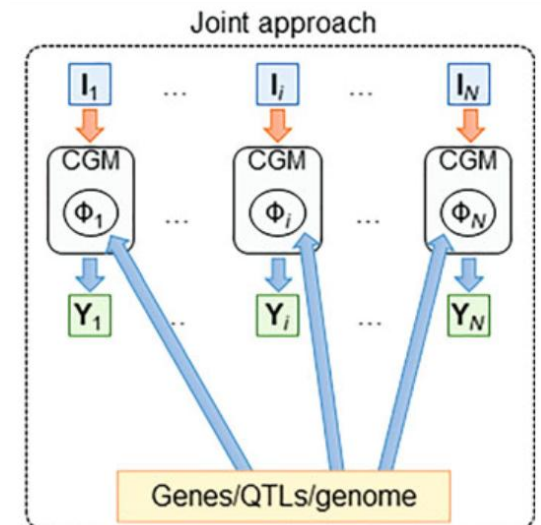
Phenological stages, environment clustering... Rincent et al. and others



Akyo Onogi (2022)

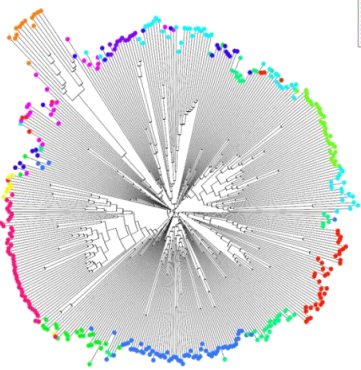


Reymond et al. 2003, 2004...



Cooper et al, Technow et al...

ML supported CGM-GP integration



- 136 accessions (only...)
From the Generation Challenge
programme Reference set



Phenoarch facility, Montpellier, France, 4 reps per genotype

Measured (dynamically)

- Number of appeared /
ligulated leaves
- Number of tillers

Measured (at harvest)

- Area of last ligulated leaf
- Plant and mainstem biomass

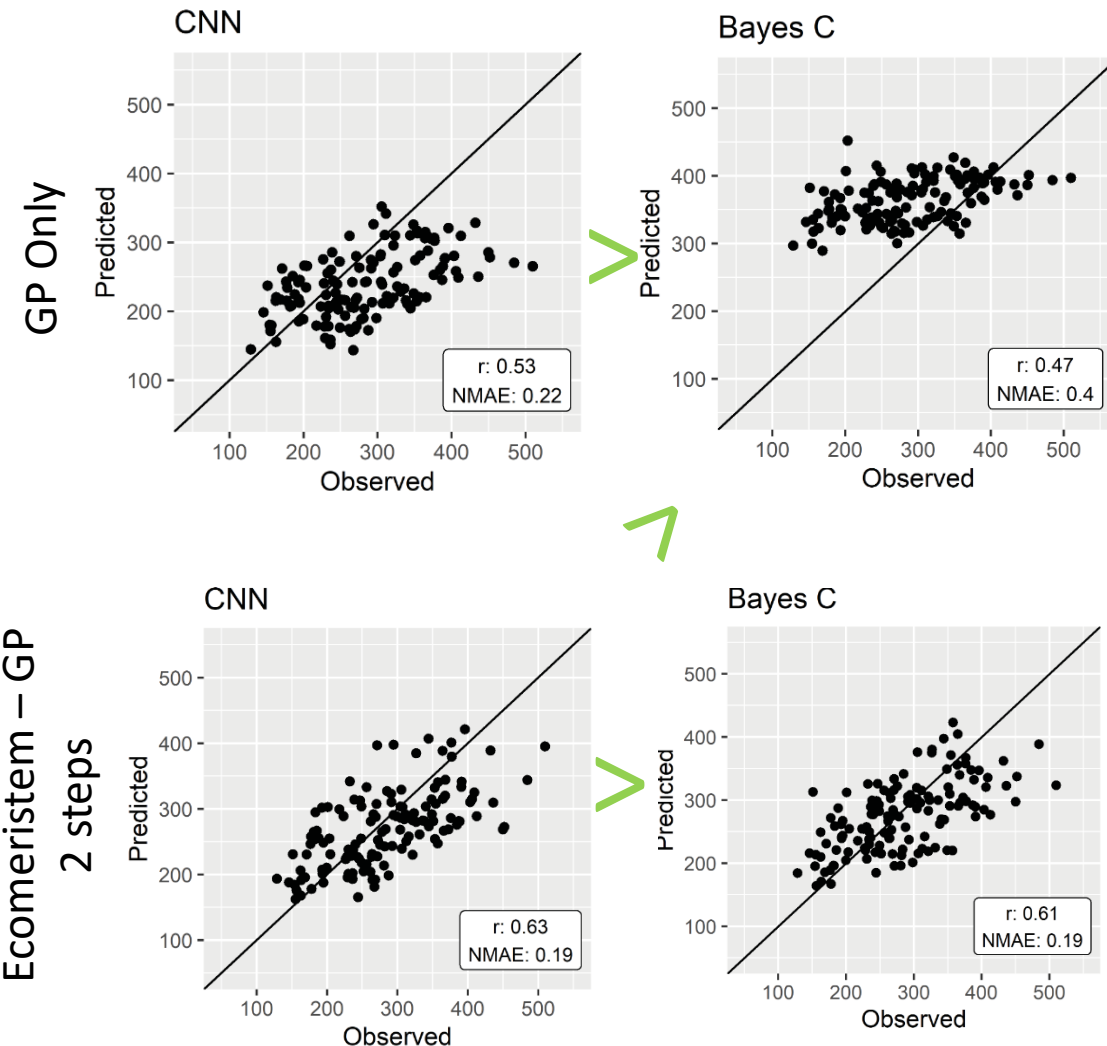
Estimated (on images)

- Plant biomass
- Plant leaf area
- Plant height
- Water use

Florian Larue (PhD)
Larue et al., 2024



Predicting Plant Aerial Biomass Ecomeristem Outputs



Coupling CGM-GP (ML/DL): Take home message

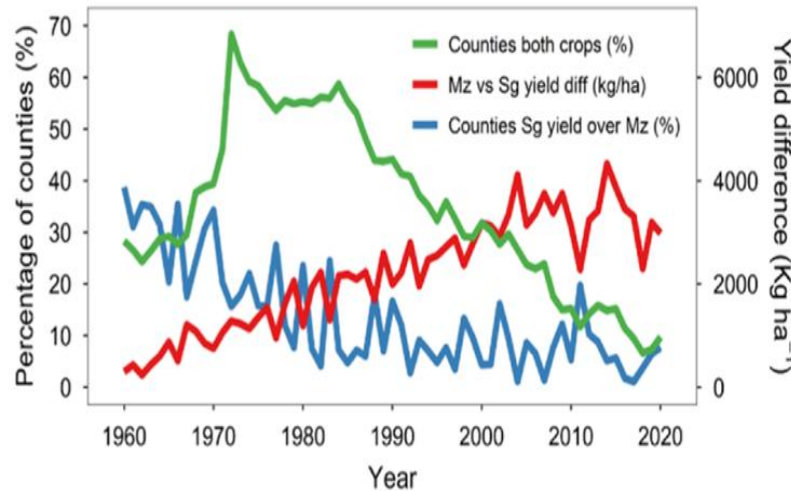
- Crop models are EXM Integrators
 - Coupling them with GP potentially allows to better predict GXEXM
- A need to rimpove the crop models to allow them to better integrate **multiple stress aspects**



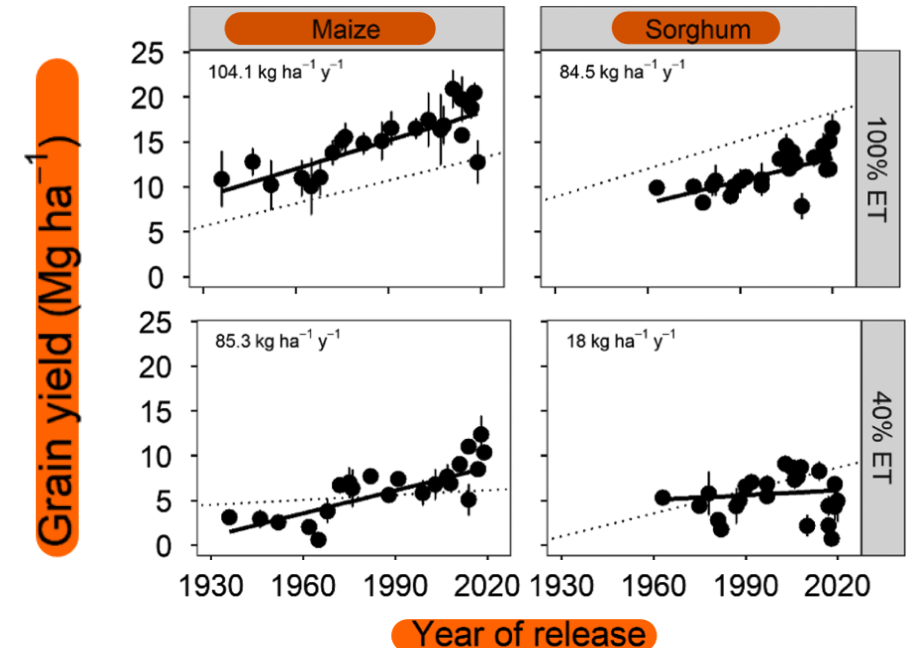
Angélique Berget
Recurrent Heat stress in Sorghum



Rotundo et al 2024



Boris Patent,
INRAE, France



Talk's outline

σ_g

Developing BCNAM resources

r

Relevance and properties of Phenomic Prediction

r

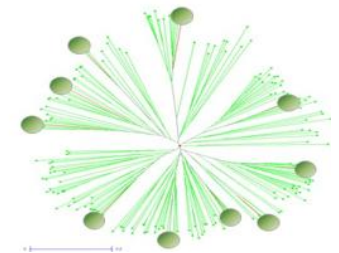
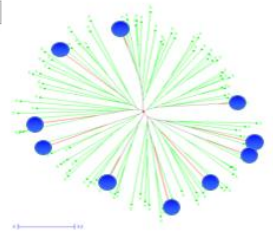
Merging CGM and Genomic prediction

σ_g

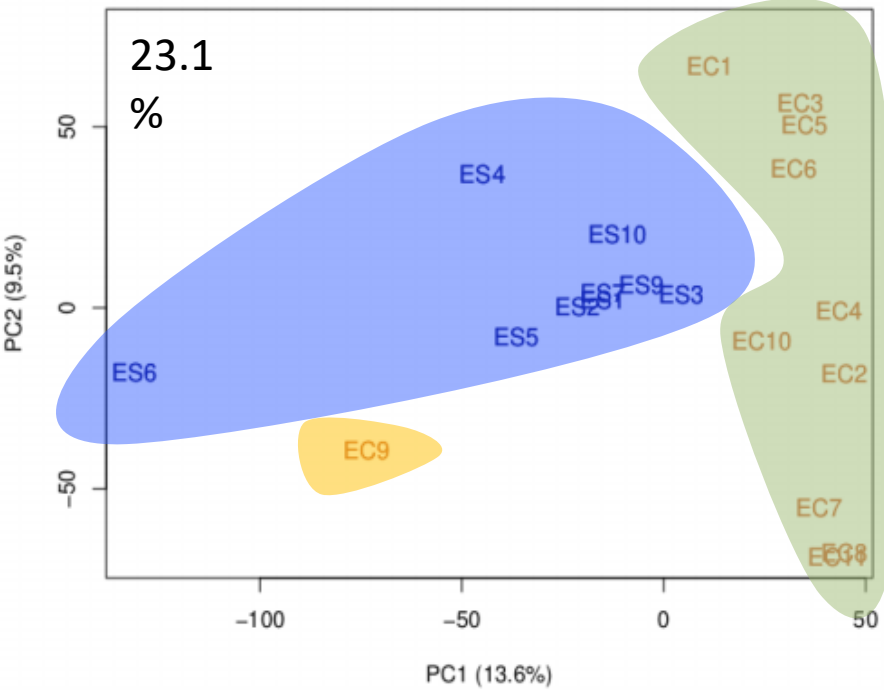
Mining the wild pool

Sorghum – Maize synergies ?

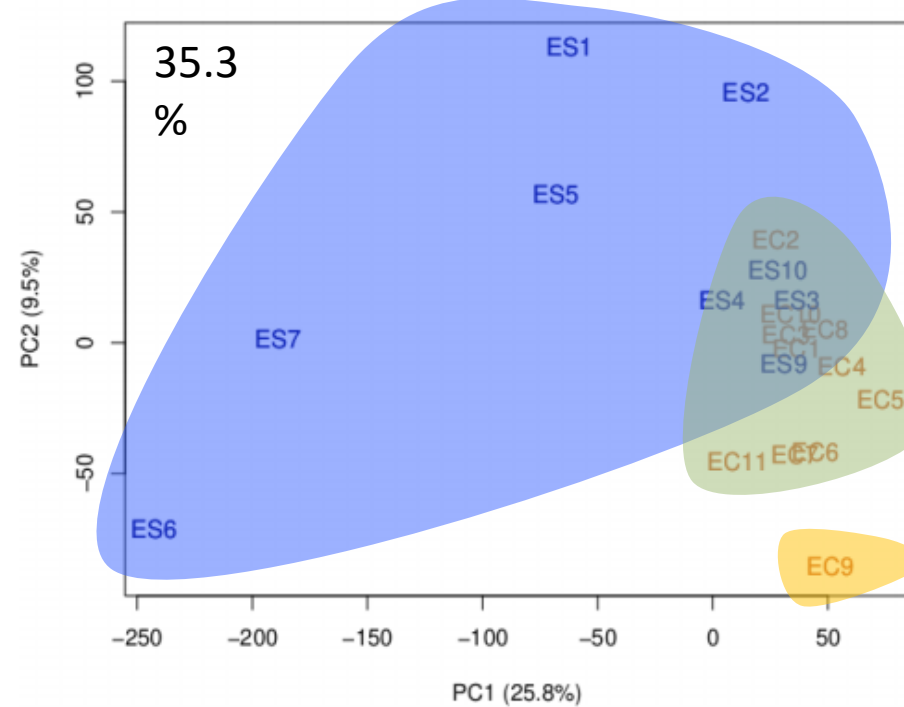
Domestication impacts differently nucleotide and transcriptome diversity structures



Nucleotide diversity



Transcriptome diversity

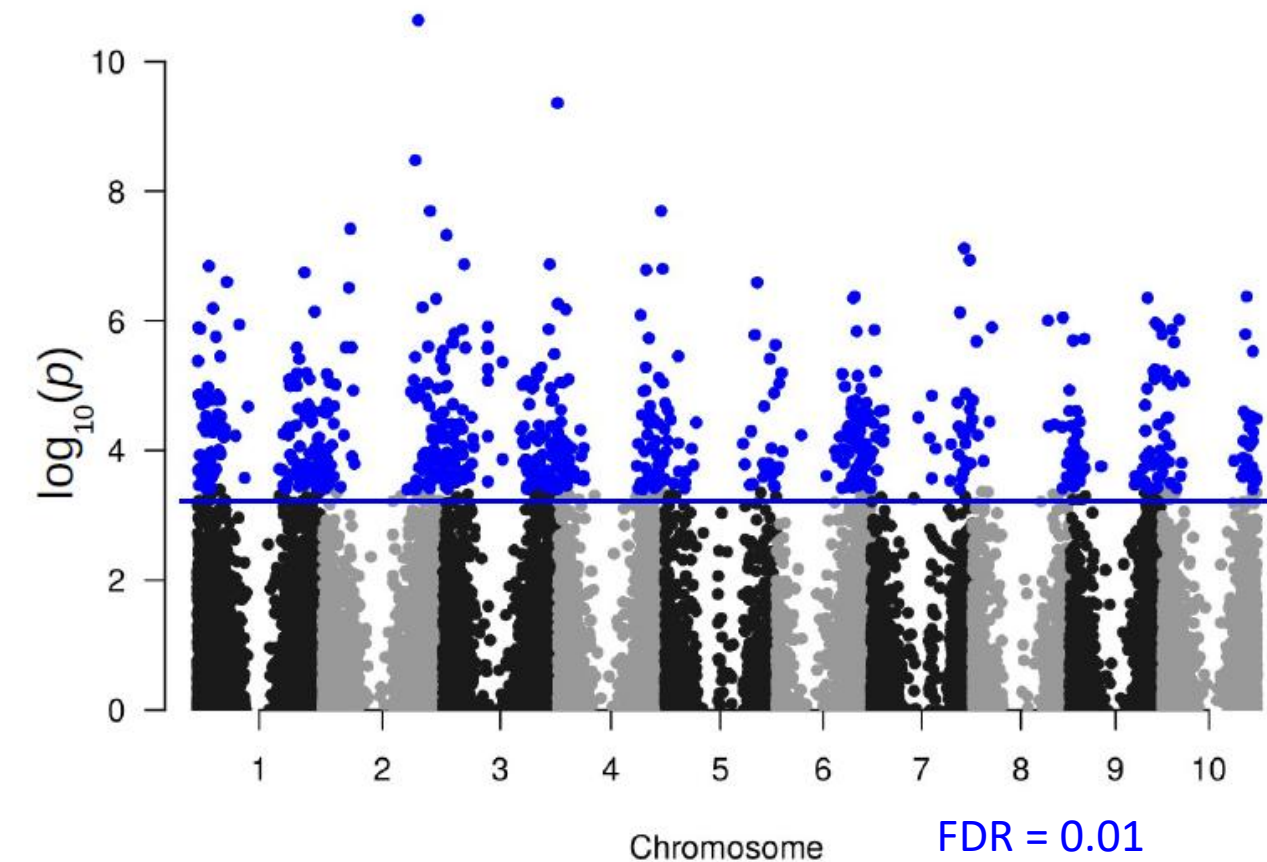


- SNP : -31 %
- Expression : -17 %
- **Gma position**

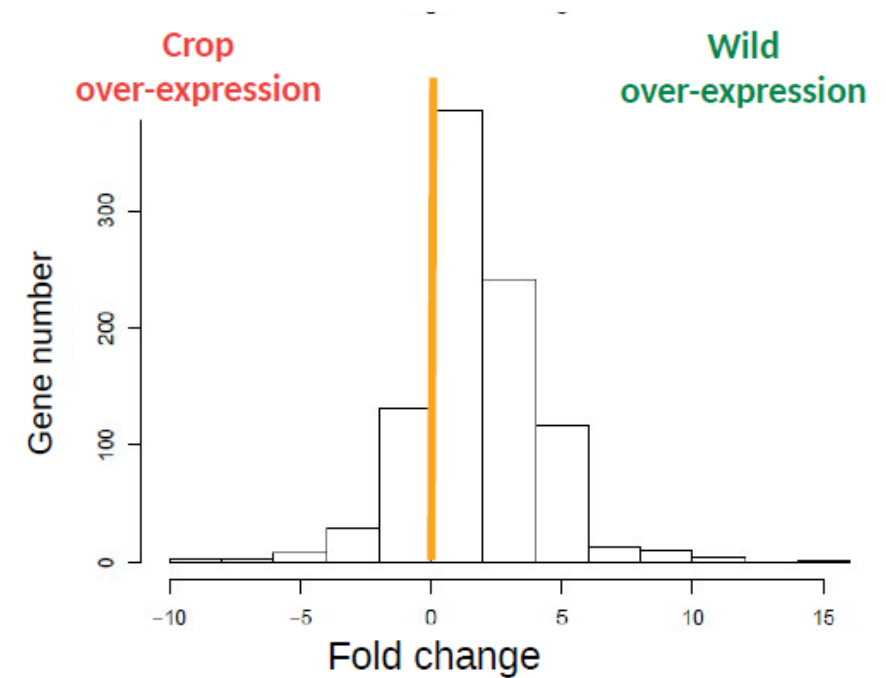


Domestication massively impact transcriptome

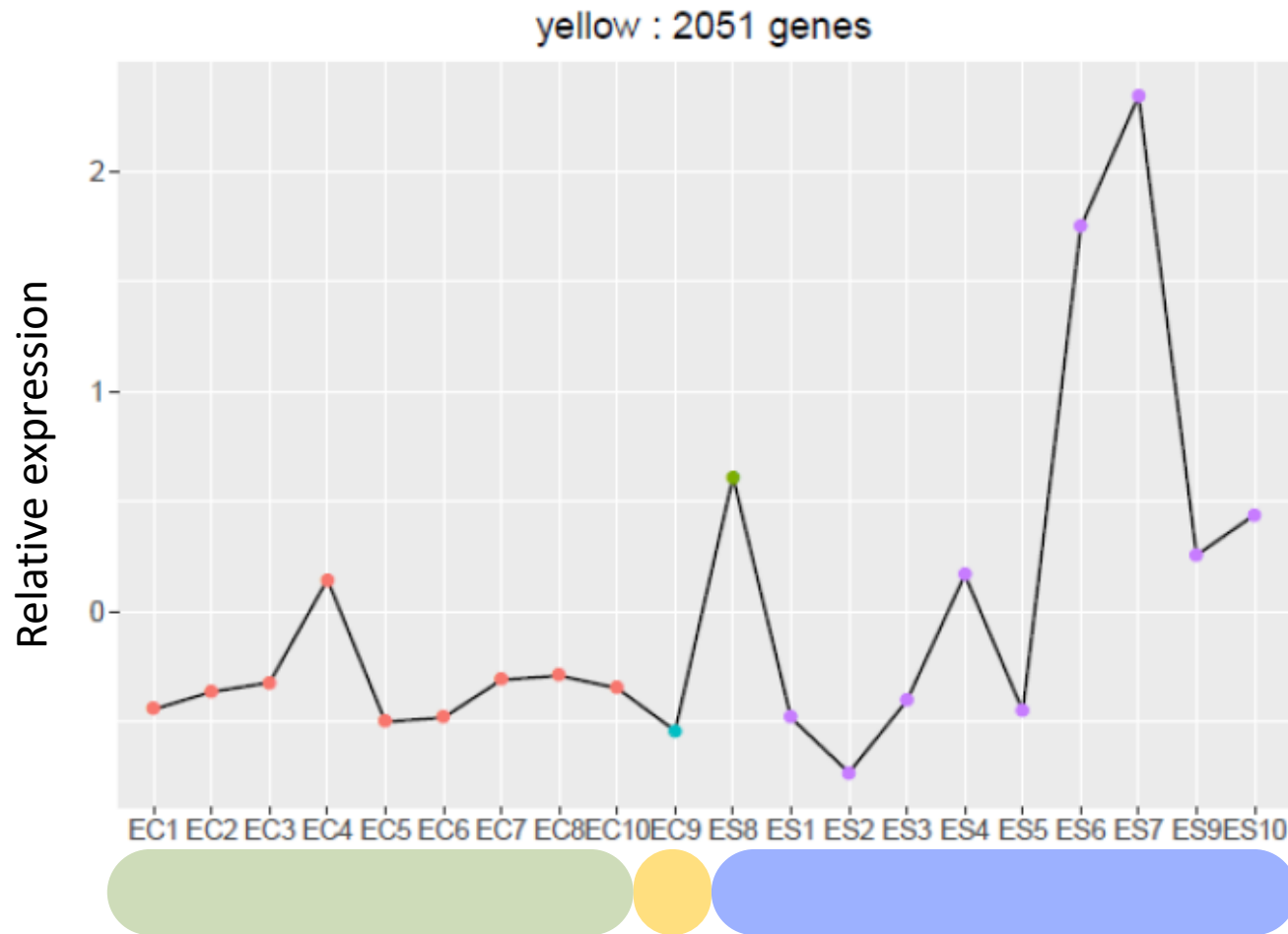
- Wild vs Crop: 949 Differentially Expressed Genes



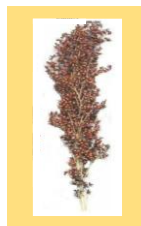
- 773 genes are Over-expressed in the wild



Gene Co-expression Networks and domestication

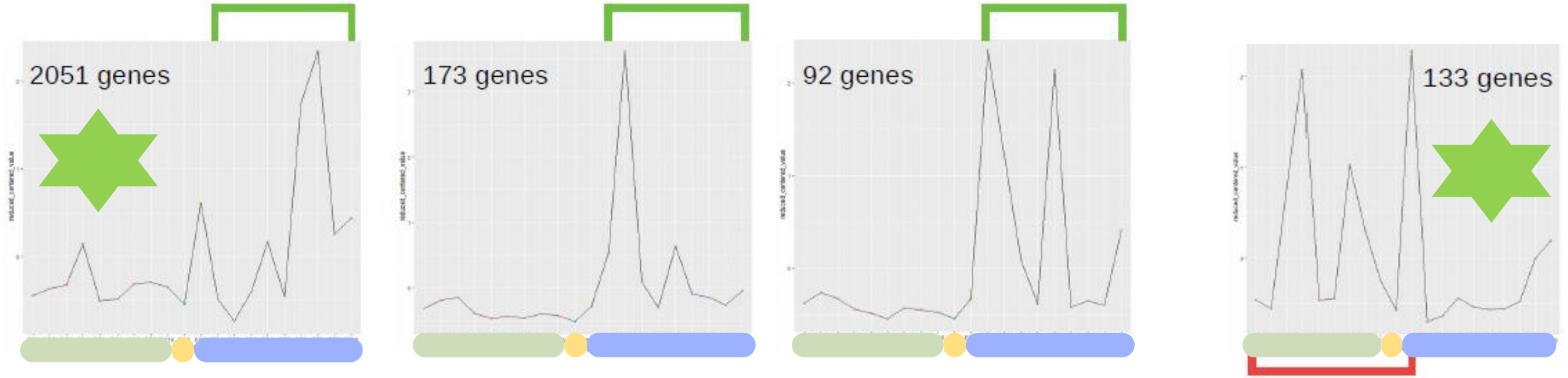


21 Modules identified



Gene Co-expression Networks and domestication

Three with over-expression and higher variability in the wild pool



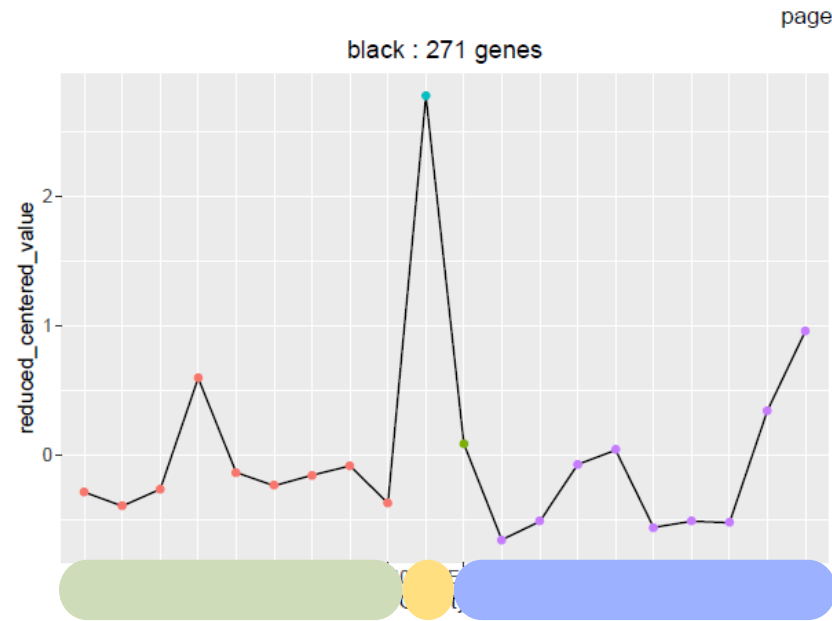
One with over-expression and higher variability in the cultivated pool



Enriched in domestication and improvement genes (Mace et al 2013)

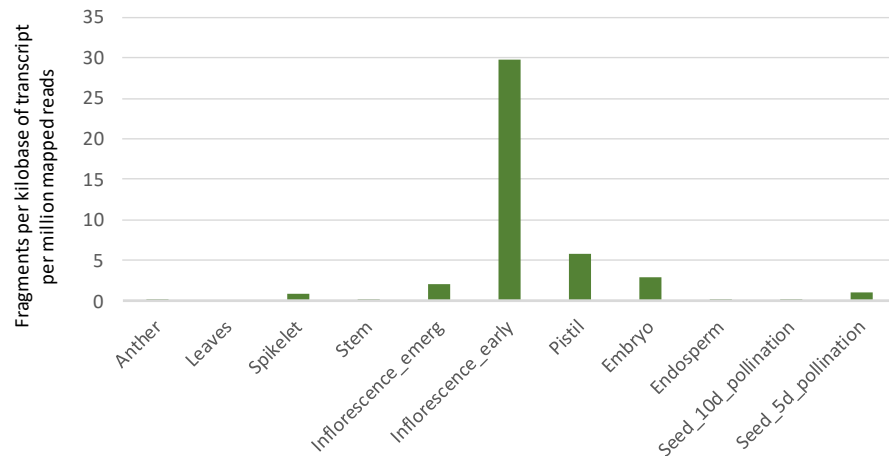


Mining the Wild to Crop continuum: Revealing key genes involved in grain size. But not only

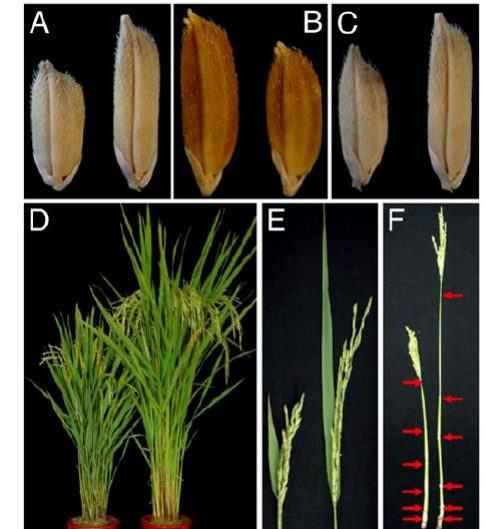
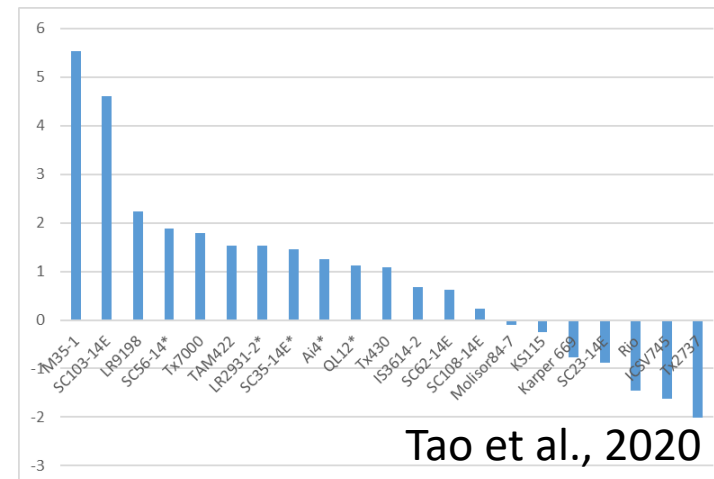


- Black module : over-expression in Guinea margaritifera (additional domestication event)
- Margaritifera are characterized as “**small vitreous grains**” (Deu and Hamon, 1994)
- Black module contains Sobic.001G341700 “SbGS3” that controls grain size in sorghum (Tao et al., 2020)

(A)



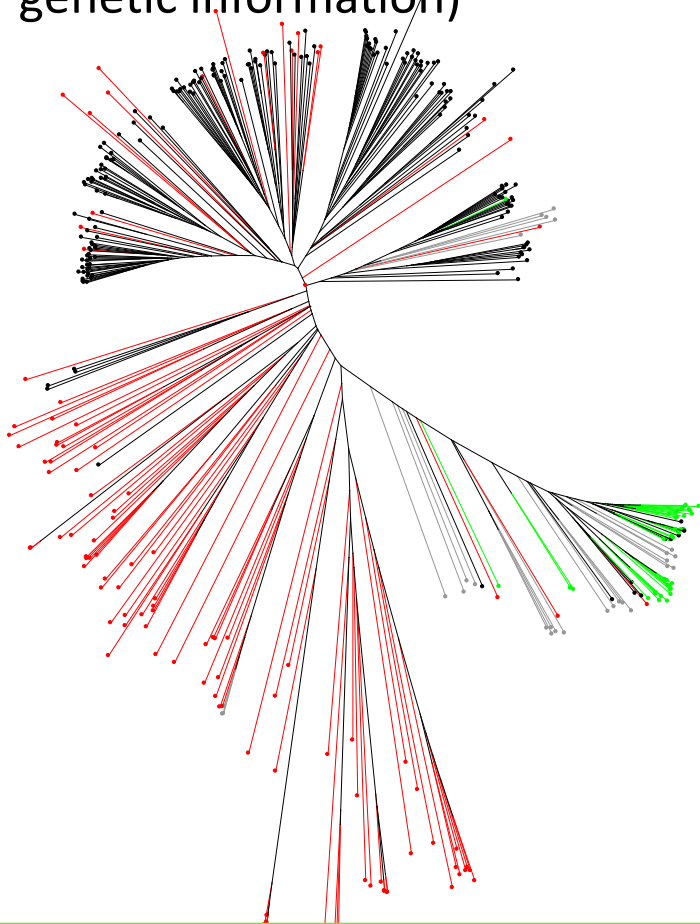
(B)



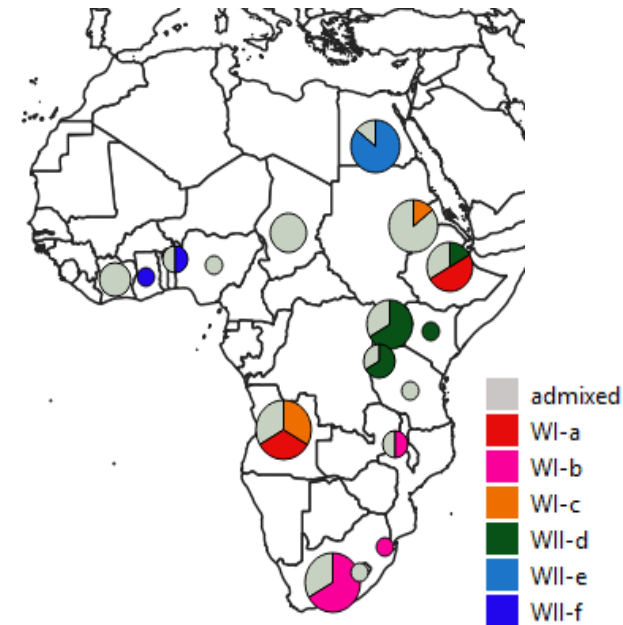
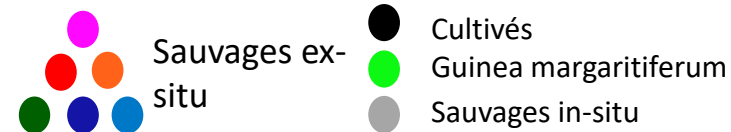
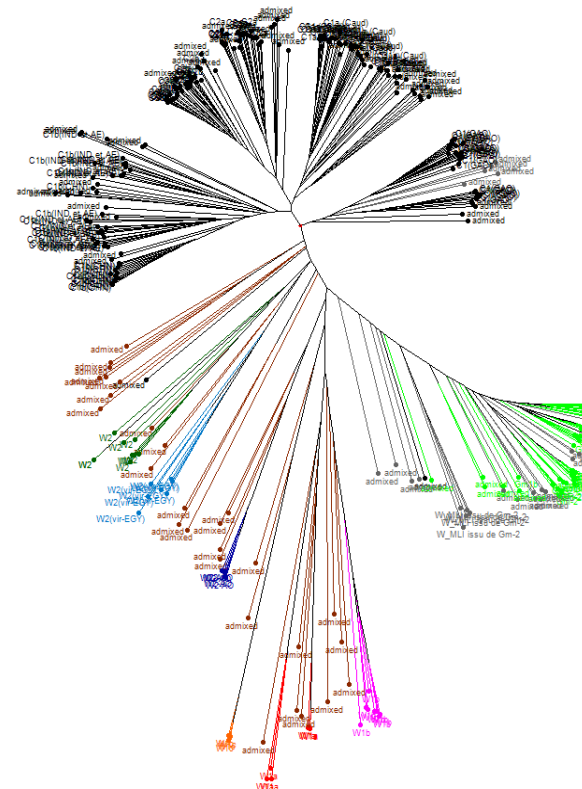
Mao et al., 2010

Refining the structure of the wild pool

- From 382 to 335 (exclusion of non pure wild based on grain size properties / genetic information)



- 335 accessions, 45 Ksnp, 92 « true » wilds



Gilabert et al,
Unpublished

Wild Genetic structure= f(Race / Geographic)

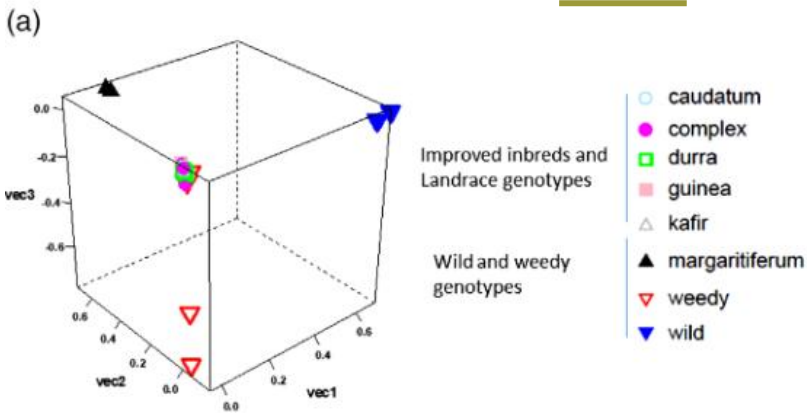
Putting wild alleles in context: A new Wild BCNAM population under development



Korotimi Thera

- 69 wild accessions multiplied in Mali
- 8 wild lines (6 genetic groups) + 1 Gm used as parent (BC1F1 available)
- Recurrent Parent : Malian Elite Variety, Photoperiod sensitive

- Complementary with the Wild BCNAM population developed in Australia (Mace et al 2020)



Donor parents	
Primary identifier	Current taxonomy
AusTRCF317030	<i>Sorghum</i> sp. (previously <i>S. versicolor</i>)
AusTRCF317961 ^{a,b}	<i>S. bicolor</i> subsp. <i>verticilliflorum</i>
IS3121 ^b	<i>S. bicolor</i> subsp. <i>verticilliflorum</i>
PI300119 ^a	<i>S. bicolor</i> subsp. <i>verticilliflorum</i>
PI330272 ^a	<i>S. bicolor</i> subsp. <i>drummondii</i>
PI525695 ^{a,b}	<i>S. bicolor</i> subsp. <i>margaritifera</i>
PI532566 ^b	<i>S. bicolor</i> subsp. <i>verticilliflorum</i>
PI535995	<i>S. bicolor</i> subsp. <i>verticilliflorum</i>
PI536008 ^b	<i>Sorghum purpureosericeum</i>
Adapted parents	
Primary identifier	Traits
Macia ^a	2-dwarf plant type, midge resistant, inbred cultivar
QL36	3-dwarf plant type, midge resistant, inbred restorer

Photoperiod insensitive

Mining the wild pool: Take home messages

- A large diversity available in the wild pool: nucleotide but also transcriptome variability
- A way to identify genes of agronomic interest (not only pathogen resistance)



Plant domestication, a unique opportunity to identify the genetic basis of adaptation

Jeffrey Ross-Ibarra, Peter L. Morrell, and Brandon S. Gaut*

2007

- A clearer description of the sorghum wild pool is available
- Wild alleles are being set-up in relevant context to allow their evaluation

Talk's outline

σ_g

Developing BCNAM resources

r

Relevance and properties of Phenomic Prediction

r

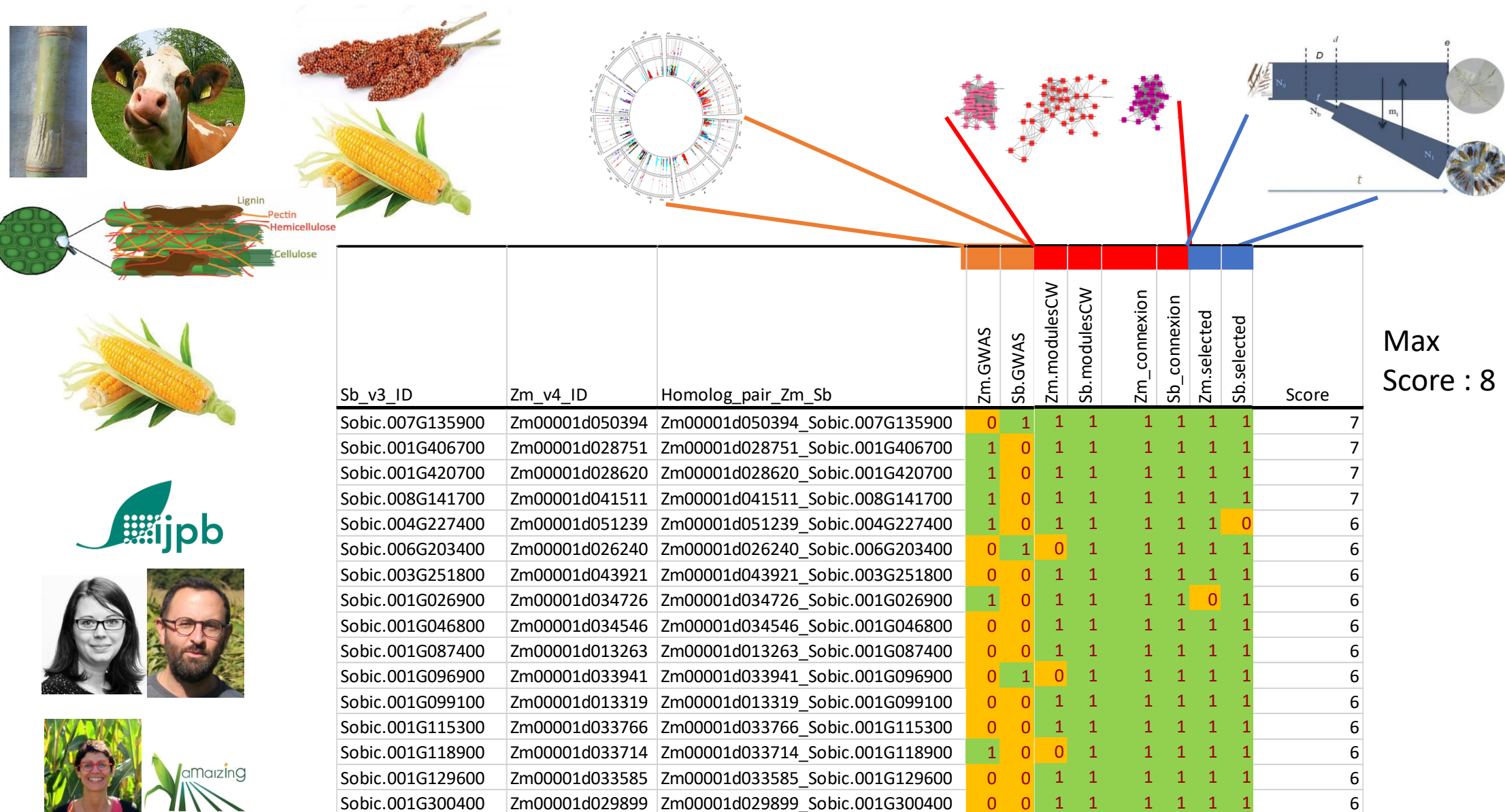
Merging CGM and Genomic prediction

σ_g

Mining the wild pool

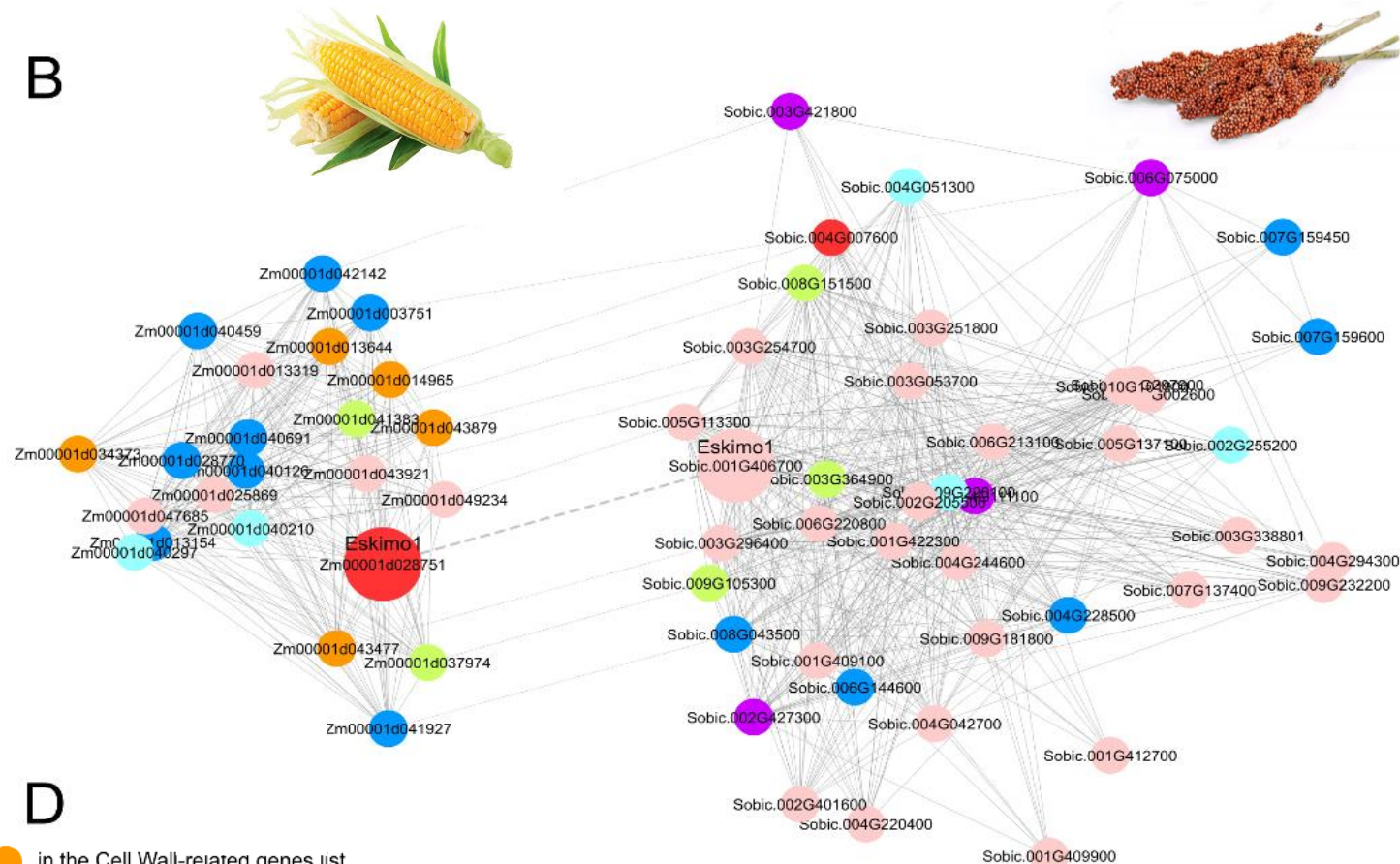
Sorghum – Maize synergies ?

Diving into the genetic determinism of cell wall composition



Score= 7 : Eskimo1, a key cell wall gene

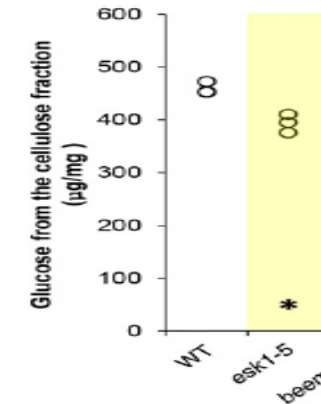
B



D

- in the Cell Wall-related genes list
- in the Cell Wall-related genes list and with a selection signature
- under a GWAS QTL
- in the Cell Wall-related genes list and under a GWAS QTL
- in the Cell Wall-related genes list and under a GWAS QTL and with a selection signature
- under a GWAS QTL and with a selection signature
- with a selection signature
- nothing

At ESKIMO1

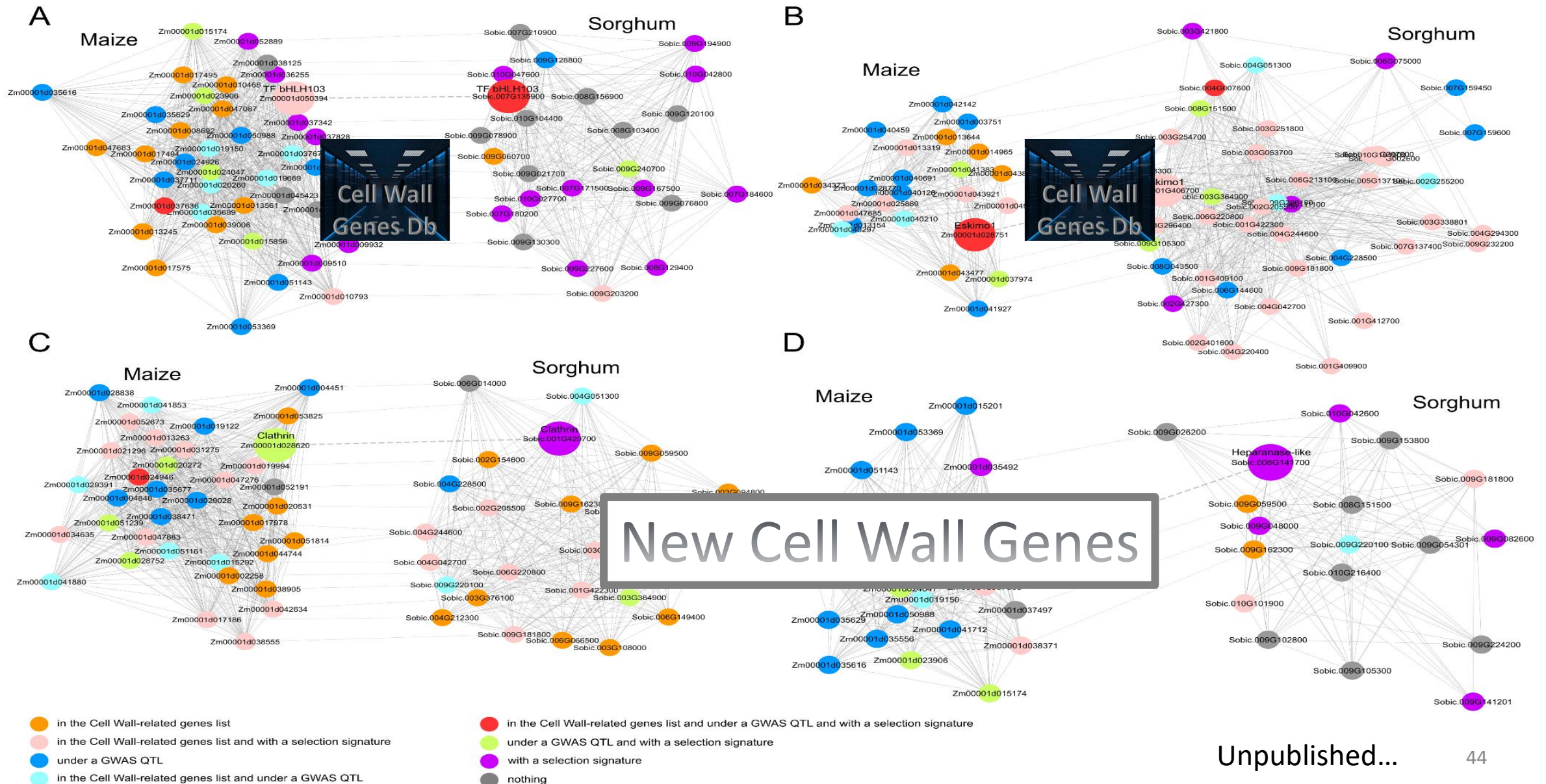


Bouchabke *et al.*, 2008; Bensussan *et al.*, 2011, Grantham *et al.*, 2017

Zm ESKIMO1 : on-going at IJPB



4 Key cell wall genes : Score = 7



Variant effect prediction across species: Improving prediction accuracy

RESEARCH

Open Access

Prediction of evolutionary constraint by genomic annotations improves functional prioritization of genomic variants in maize

Guillaume P. Ramstein^{1,2*} and Edward S. Buckler^{2,3}



- Nucleotide conservation across angiosperms, as a proxy for fitness effect of mutations
- Deep learning methods mobilized to characterize nucleotide conservation
- « Improves genomic prediction for fitness related traits such as grain yield, in elite maize panels, by stringent prioritization of fewer than 1% of single-site variants »



Received: 26 July 2024 | Accepted: 14 November 2024

DOI: 10.1002/tpg2.20553

ORIGINAL ARTICLE

Special Section: Tribute to Ron Phillips: Crop Genetics, Genomics and Biotechnology

Improvement in genomic prediction of maize with prior gene ontology information depends on traits and environmental conditions

Baber Ali¹ | Tristan Mary-Huard^{1,2} | Alain Charcosset¹ | Laurence Moreau¹ | Renaud Rincent¹

PNAS

RESEARCH ARTICLE

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

OPEN ACCESS



DNA language models are powerful predictors of genome-wide variant effects

Gonzalo Benegas^a, Sanjit Singh Batra^b, and Yun S. Song^{b,c,d,1}

nature communications



Article

<https://doi.org/10.1038/s41467-024-47744-C>

Deep learning the *cis*-regulatory code for gene expression in selected model plants

Received: 28 April 2023

Fritz Forbang Peleke^{1,7}, Simon Maria Zumkeller^{2,3,7}, Mehmet Gültas⁴, Armin Schmitt^{5,6} & Jędrzej Szymański^{1,2,3}

Take home messages

$$\Delta G = \frac{i \quad r \quad \sigma_g}{T}$$

- Developing populations allowing to assess alleles effects and allow their use in breeding programs
 - Cultivated and wild resources (on-going)
- Ability to improve prediction accuracy
 - CGM – GP integration
 - GP supported by functional annotations (in Maize at least)
- Crop model optimization
- Gene / motifs (CRE) / variant prediction effects
- Not only an AI problem, key importance of ability to connect different information in a meaningful way



Understanding still a lever to better predict

Thanks for your attention
Questions !