Génétique
Quantitative
et Évolution
Le Moulon

MineLandDiv
Mining allelic diversity of maize landraces
for tolerance to abiotic and biotic stresses

Predicting seed composition and agronomic traits in maize genetic resources using phenomic selection based on near-infrared spectra

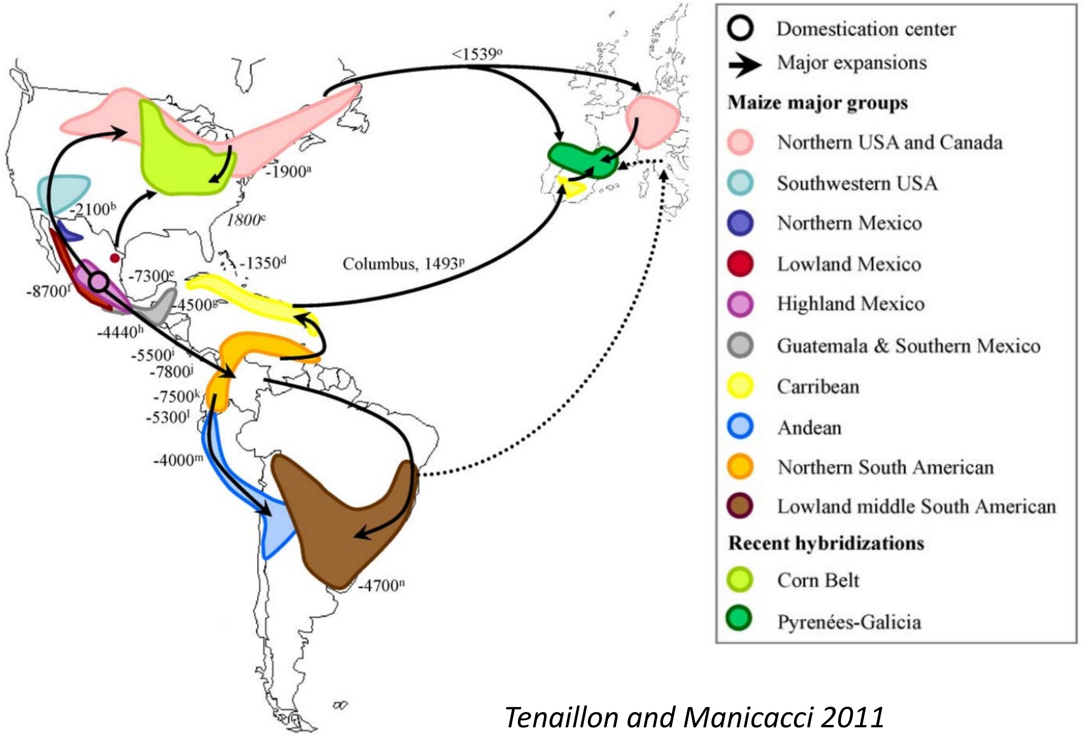
Stéphane Nicolas, Sarah Ben-Sadoun Renaud Rincent,
Laurence Moreau, Alain Charcosset, Agustin Galaretto,
Antoine Roux, Brigitte Gouesnard

stephane.nicolas@inrae.fr
GQE- Le Moulon, INRAE

PRO-GRACE/EMPHASIS SYMPOSIUM
27-28 June 2024

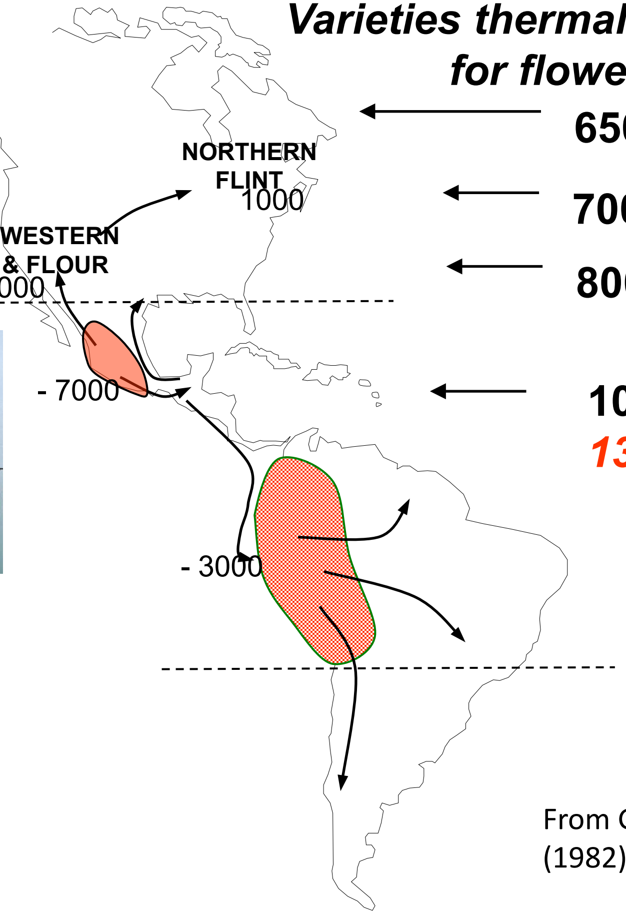
INTRODUCTION: Maize genetic diversity was shaped by domestication, expansion and adaptation to ≠ environments

Domestication 9000 years ago in Mexico => Expansion => Recent introductions to Europe



Tenaillon and Manicacci 2011
Tenaillon and Charcosset 2011

Southwards and northwards migration accompanied by drastic change in phenology



Varieties thermal requirement for flowering:

650 GDD

700 GDD

800 GDD

1000 GDD short days

1300 GDD long days

From Gerrish (1982)

Landraces were well adapted to different local agro-climatic conditions
=> Original and promising alleles for breeding to face climatic change / low input agriculture?

INTRODUCTION: Genetic diversity of current elite germplasm was shaped by the transition from the landraces (OPV) to F1 hybrids and modern breeding

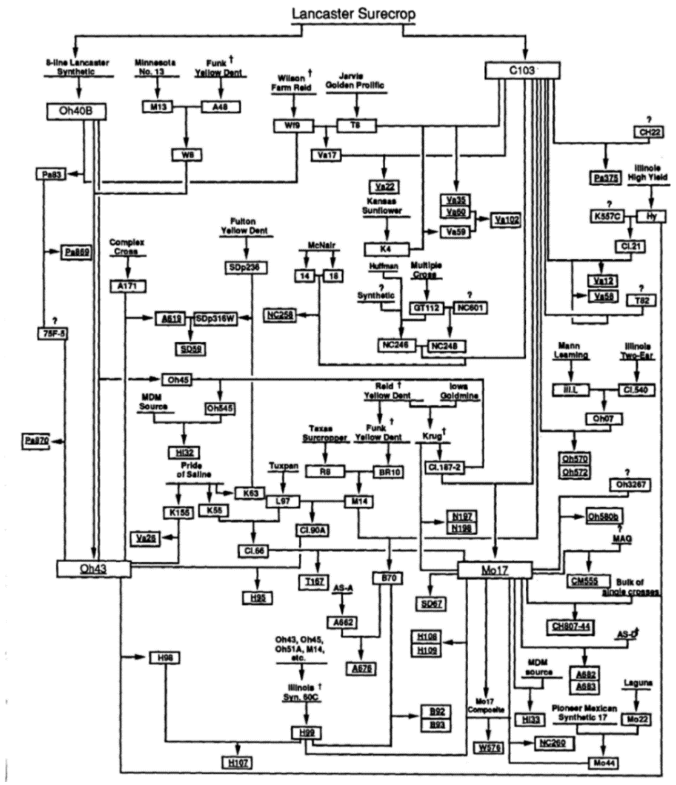
Landraces = Open Pollinated Varieties

~1930 (USA) / ~1950 (EUR)

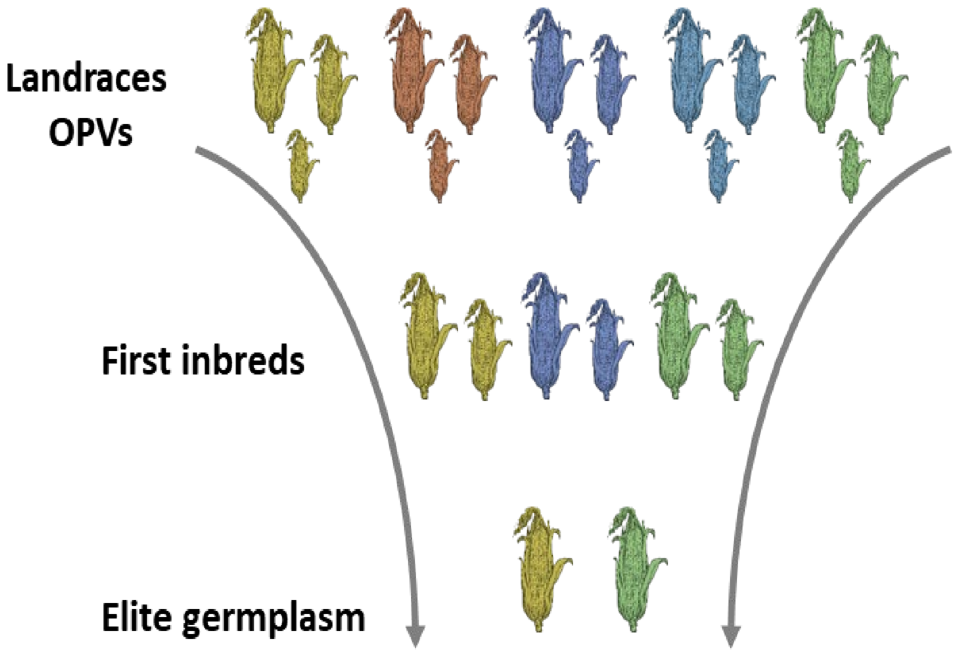
First Cycle inbred lines and F1 hybrids

Reciprocal recurrent selection

Elite breeding germplasm



Lancaster Surecrop derived lines
(Gerdes et Tracy, 1993)

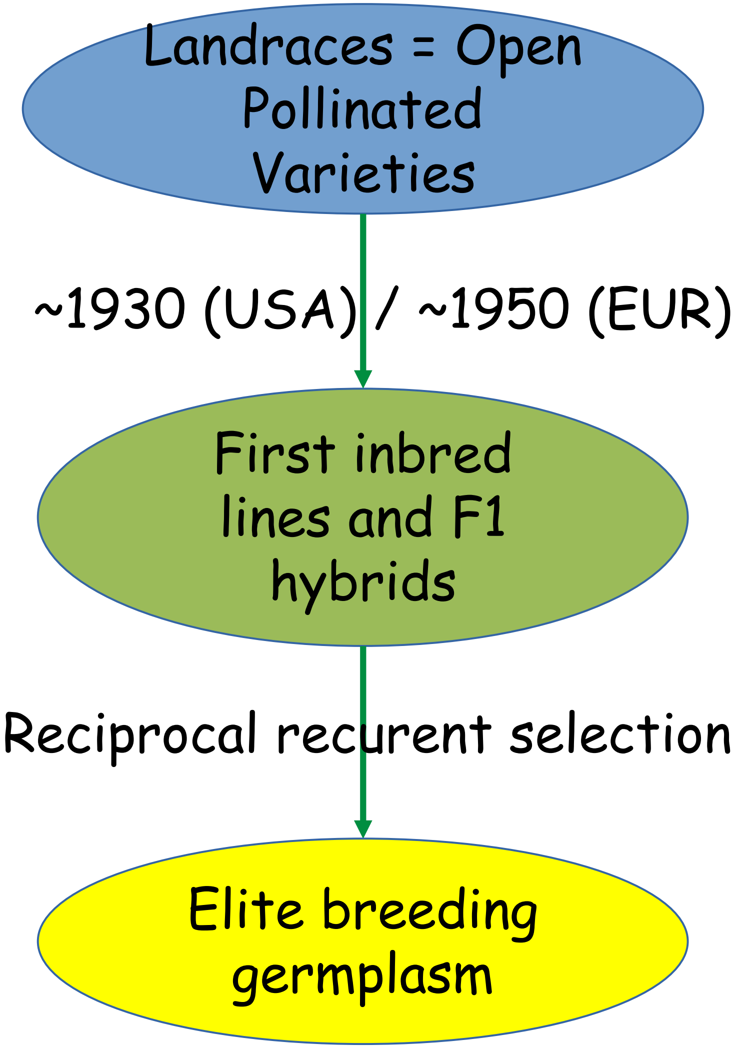


Few landraces contributed to modern breeding germplasm and recurrent selection erode progressively genetic diversity

=> Loss of allelic diversity

=> Untapped source of genetic diversity for prebreeding and new traits required for low-input agriculture an for facing climate change?

INTRODUCTION: Fortunatly, “Genetic Resources” collections have been put in place in parallel to the replacement of landraces



Landraces = Open Pollinated Varieties

~1930 (USA) / ~1950 (EUR)

First inbred lines and F1 hybrids

Reciprocal recurent selection

Elite breeding germplasm

Large number of landraces (and inbred lines) maintained ex-situ in genebank (seeds) and in-situ (farmers):

	CIMMYT	USDA	INRAE	World
Landraces	26.956	14.400	1.600*	>135.000
Inbred lines	643	2.780	3.000**	

* CRB Gamet, Montpellier (resp. M. Vincent)
** CRB lignées Saint Martin de Hinx (resp. C. Palaffre)

But limited use in breeding program and by farmers due to lack of characterization, gap in performance, and large within genetic diversity for landraces

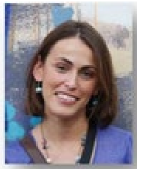
=> Need low cost approach (genotyping, NIRs) for characterizing landraces (and inbred lines) for traits including seeds composition in genetic resources for addressing challenge of low input agriculture, new uses and climate change

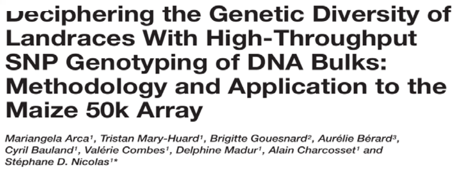
How to retrieve at low cost the genetic values of maize landraces for different traits including seed composition in different environnements?

Part I: Predict traits of landraces in genebank with genomic selection

Part I: Predict traits of landraces in genebank with genomic selection

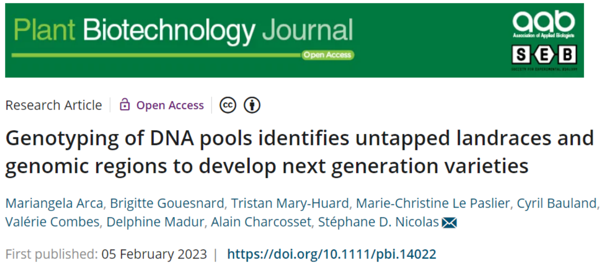
Development of a DNA bulk method based on 50K array to genotype landraces (allelic frequencies) because large within genetic diversity and identify promising source of diversity (Arca et al., 2021, Arca et al., 2023, Diaw et al., 2021, Balconi, Galaretto et al., 2024)





Deciphering the Genetic Diversity of Landraces With High-Throughput SNP Genotyping of DNA Bulks: Methodology and Application to the Maize 50k Array

Mariangela Arca[1], Tristan Mary-Huard[1], Brigitte Gouesnard[2], Aurélie Bérard[3], Cyril Bauland[1], Valérie Combes[1], Delphine Madur[1], Alain Charcosset[1] and Stéphane D. Nicolas[1]*



Plant Biotechnology Journal

Open Access

Research Article | Open Access

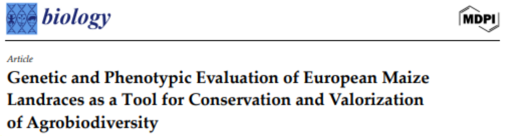
Genotyping of DNA pools identifies untapped landraces and genomic regions to develop next generation varieties

Mariangela Arca, Brigitte Gouesnard, Tristan Mary-Huard, Marie-Christine Le Paslier, Cyril Bauland, Valérie Combes, Delphine Madur, Alain Charcosset, Stéphane D. Nicolas

First published: 05 February 2023 | https://doi.org/10.1111/pbi.14022

European Cooperative Programme for Plant Genetic Resources

ECP/GR

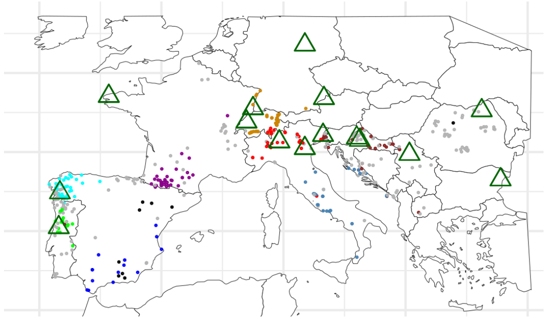


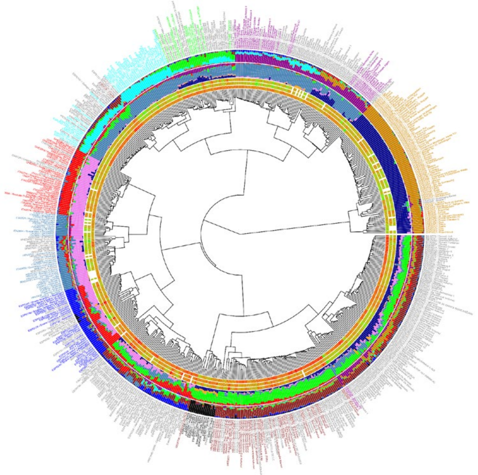
biology

MDPI

Article

Genetic and Phenotypic Evaluation of European Maize Landraces as a Tool for Conservation and Valorization of Agrobiodiversity





Development of a genomic selection approach based to predict traits in landraces (Ben Sadoun, Galaretto et al., in prep)

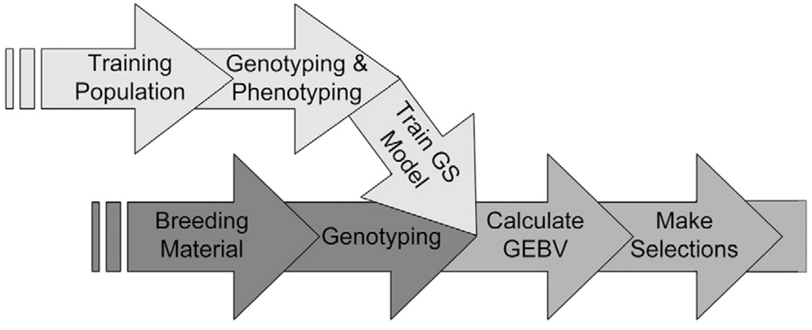


Figure 1. Diagram of genomic selection (GS) processes starting from the training

Heffner et al. 2009

Adj. means of the reference individuals

Polygenic random effect of all indiv.

Kinship

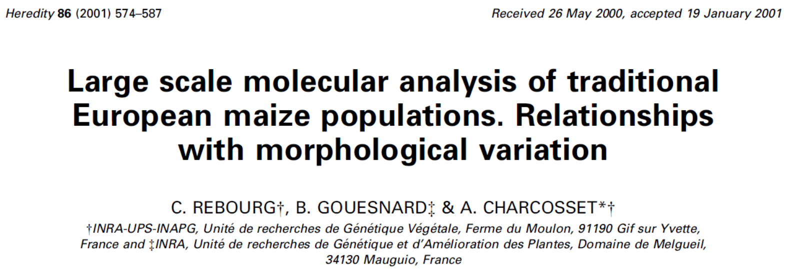
$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} \qquad var(\boldsymbol{u}) = \boldsymbol{A}\sigma_g^2$$

- Kinship estimated with genotyping data (allelic frequency)
- Allows predicting new individuals (unphenotyped)

1,623 landraces currently genotyped with 50K using DNA bulk => Database to identify promising sources of genetic diversity

Part I: Predict traits of landraces in genebank with genomic selection

Combine landraces phenotypic with genotypic data to predict traits in landraces



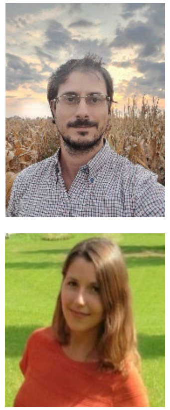
Heredity **86** (2001) 574–587 *Received 26 May 2000, accepted 19 January 2001*

Large scale molecular analysis of traditional European maize populations. Relationships with morphological variation

C. REBOURG†, B. GOUESNARD‡ & A. CHARCOSSET*†

†INRA-UPS-INAPG, Unité de recherches de Génétique Végétale, Ferme du Moulon, 91190 Gif sur Yvette, France and ‡INRA, Unité de recherches de Génétique et d'Amélioration des Plantes, Domaine de Melgueil, 34130 Mauguio, France

(Rebourg *et al.*, 2001)



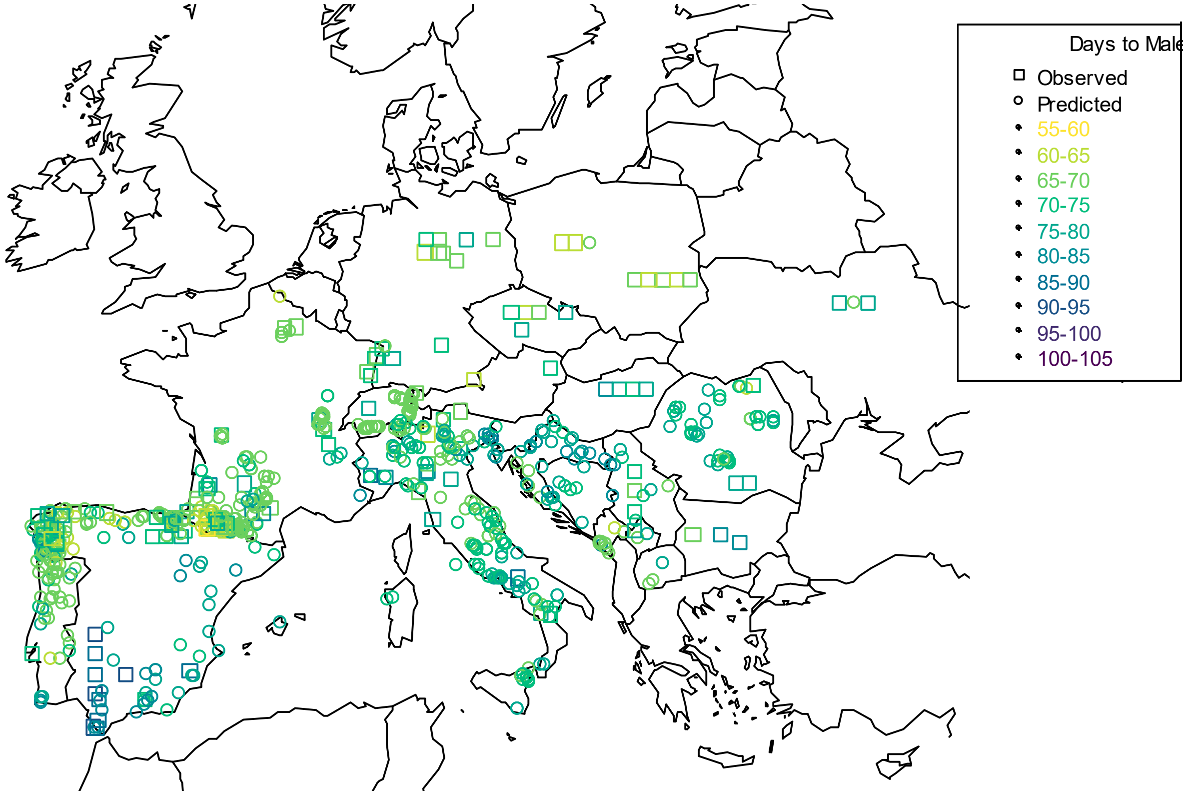
The training set of the GBLUP model

171 landraces from Europe (129) and America (42) evaluated in 1998 at two locations for flowering time (Moulon, Mauguio) with 2 replicates
Genotyped with 50K

Genomic prediction on 1047 temperate landraces genotyped with 50K using GBPLUP

$$y = Xb + Za + \varepsilon$$

Map variation of flowering time variation accros Europe



How to predict seeds composition of landrace accessions in genebank ?

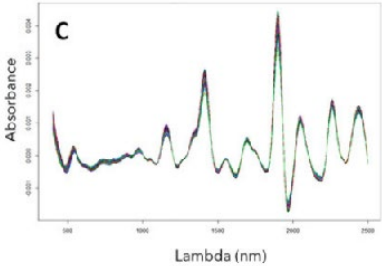
Part II: Predict seed composition of landraces with NIRs and extend this characterization by genomic selection

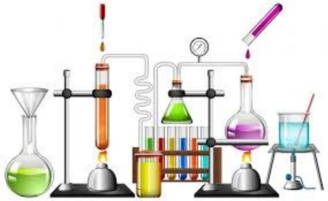
Part II: Predict seed composition of landraces with NIRs

NIRs, a low cost and non destructive approach classicaly use in breeding program to evaluate seeds biochemical composition (or digestibility) using NIRs on seeds (or silage) harvested. This biochemichal characterization is based on a model trained on a set of seed (silage) samples that are analyzed both with NIRs and with wet lab

Training set

NIRs Infratec NOVA

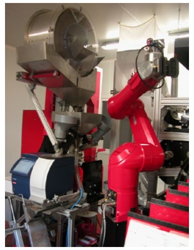


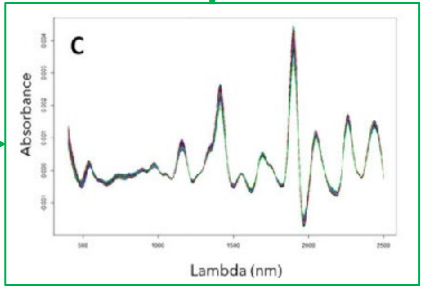


Calibrate NIRs equation with Partial Least Square model (PLS, Neutral network) Y ~ NIR wavelength
Y = Protein content,…

Sample ID	Starch content (%)	Oil content (%)	Water Content (%)	Protein Content (%)	TKW (g)
1237	73,0	4,0	16,3	9,5	227,3
1109	72,6	4,6	15,4	8,6	262,3
1219	72,5	4,3	15,2	8,7	374,7
1118	71,8	4,7	15,1	9,4	311,3
1144	72,4	4,7	15,0	8,2	280,9
1222	71,4	4,8	14,9	9,4	364,1
1246	71,7	4,8	15,8	9,3	313,9
1239	71,9	4,5	15,3	9,0	331,4
1147	71,4	4,6	15,2	9,7	365,0
1152	73,6	4,5	16,3	7,6	293,4

Samples to predict

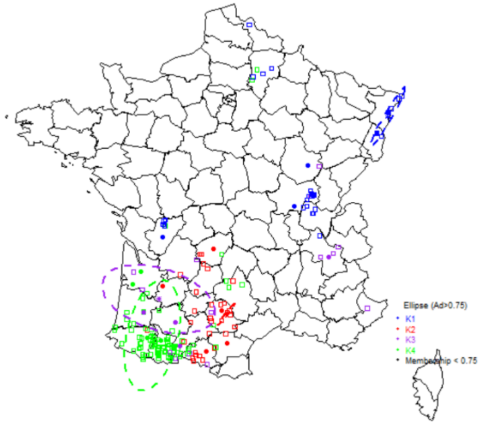




=> DEVELOPMENT OF AUTOMATED HIGH TROUGPUT PORT HARVESTED DEVICES FOR NIRs, Seeds morphology (UE DIASCOPE, F. Meunier) / AGAP Institute (M. Ecarnot)

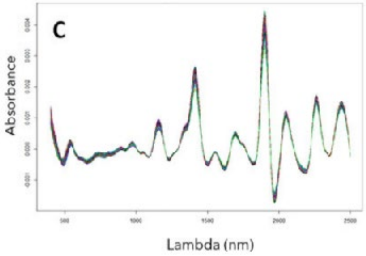
Part II: GWAS on DNA bulk identified genomic regions involved in agronomic trait and seed quality traits based on NIRs

263 french landraces including genotyped with 30,698 SNPs



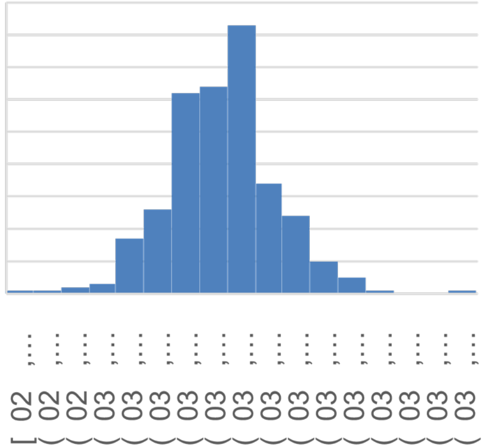
Ben Sadoun, Galaretto et al., in prep

NIRs evaluation of seeds of 263 landraces harvested in two trials for evaluating seed composition + agronomi traits (Flowering time, TKW…)



NIRs equation

Linoleic oil content (% DM)



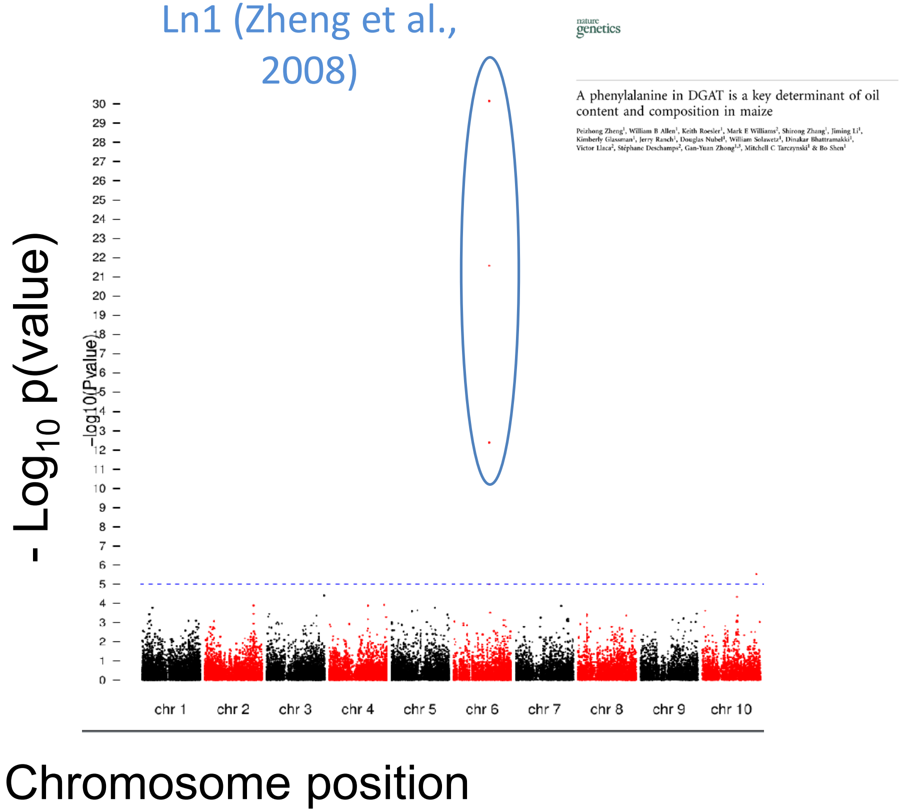
GWAS on seed quality traits based on NIRs prediction

Mixed model:

$Y = X\beta + Zu + e$

$u \sim N(0, K\sigma_g)$

Linoleic oil content (% DM)



A phenylalanine in DGAT is a key determinant of oil content and composition in maize

Peizhong Zheng[1], William B Allen[1], Keith Roesler[1], Mark E Williams[2], Shirong Zhang[1], Jiming Li[1], Kimberly Glassman[1], Jerry Ranch[1], Douglas Nubel[1], William Solawetz[1], Dinakar Bhattramakki[1], Victor Llaca[2], Stéphane Deschamps[2], Gan-Yuan Zhong[1,3], Mitchell C Tarczynski[1] & Bo Shen[1]

GWAS identified QTLs that pointed to know genes involved in seed composition as exemplified by Ln1 genes for linoleic oil content variation or Krn4 for kernel row number

Part II: Predict seed composition of landraces with NIRs and extend this characterization by genomic selection

Combine landraces phenotyped with seed composition and genotypic data to predict seed composition of unphenotyped landraces based on their genotyping

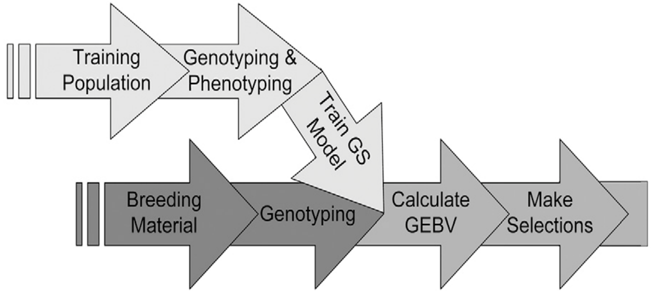
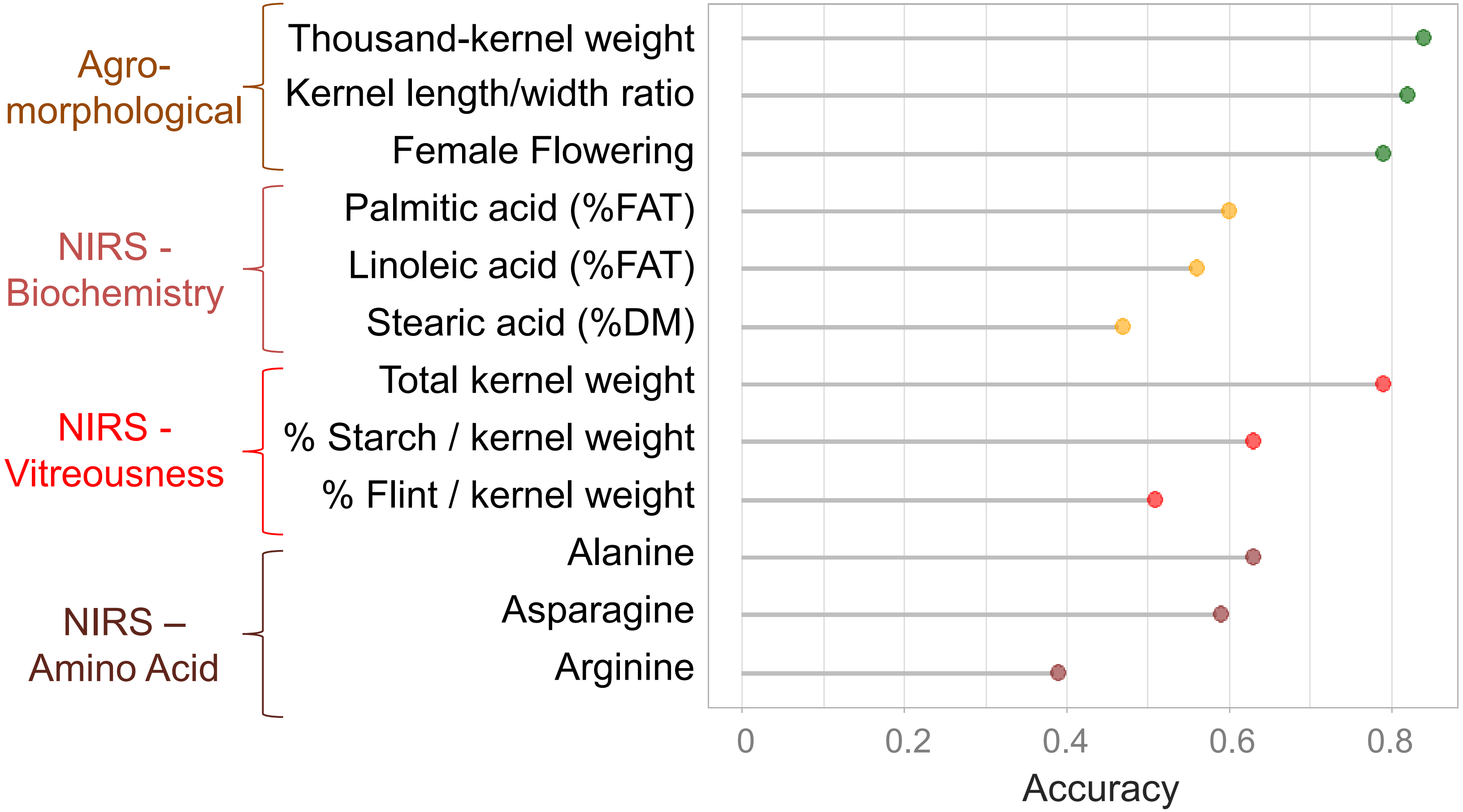


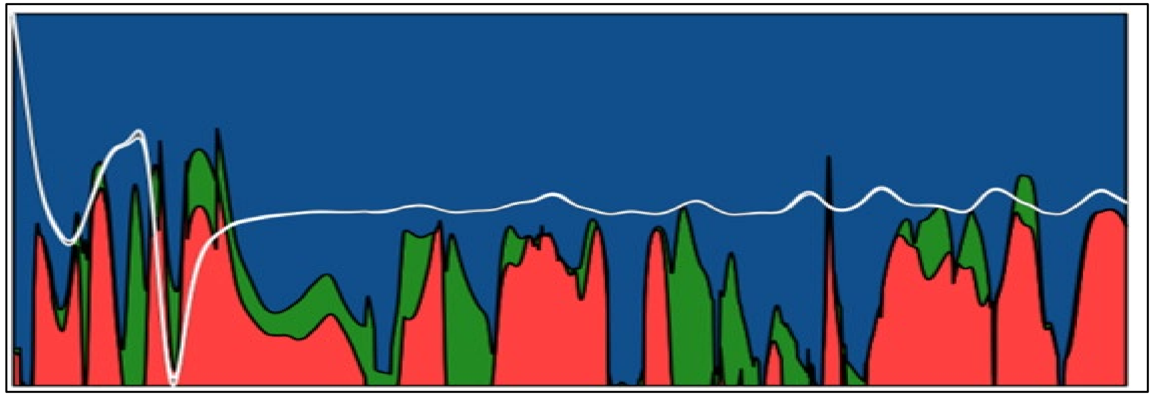
Figure 1. Diagram of genomic selection (GS) processes starting from the training

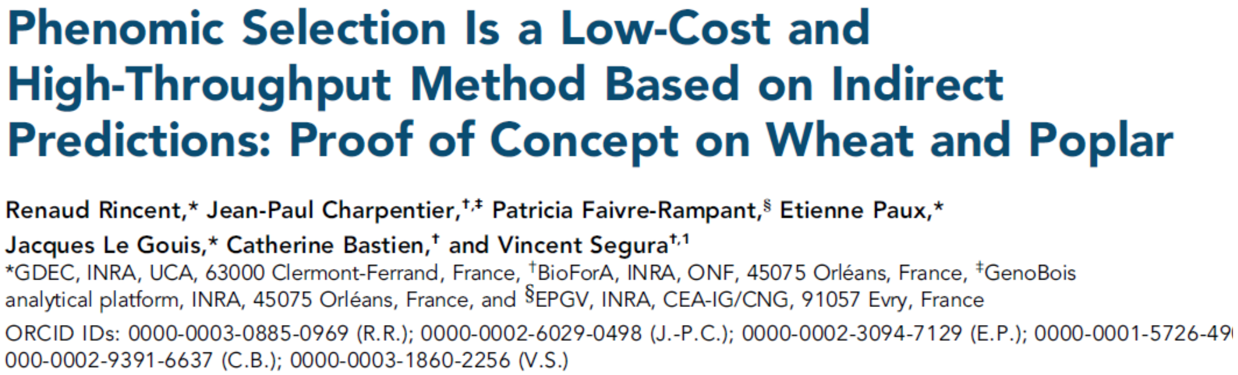
$$y = X\beta + Zu + e$$



High accuracy for predicting different agro-morphological traits and some traits related to grain quality using NIRs => A way to characterize landraces maintained in genebanks

Part III: Phenomic selection, a new way of using NIR to predict traits of accessions in the genebank



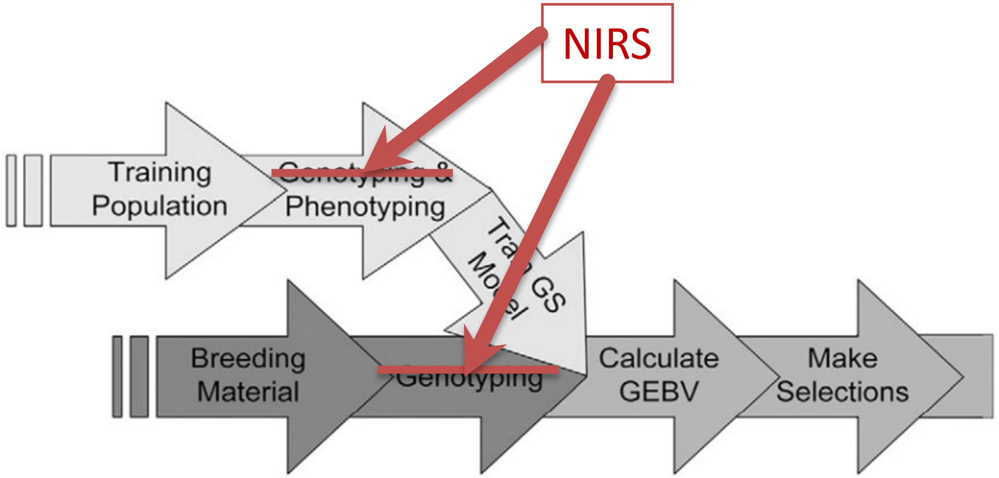


Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar

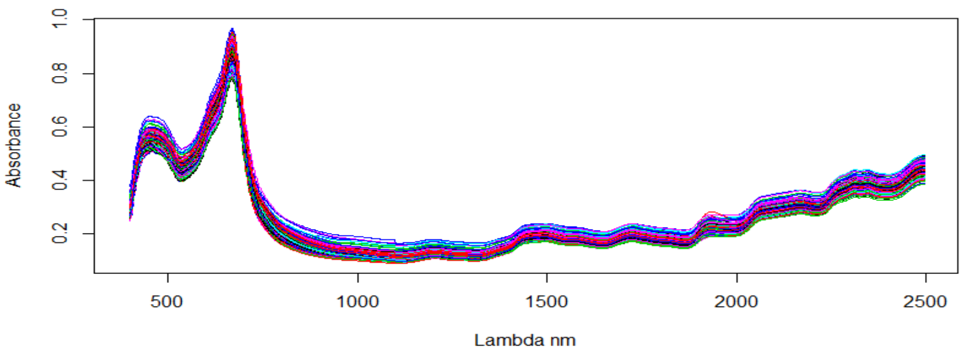
Renaud Rincent,* Jean-Paul Charpentier,†,‡ Patricia Faivre-Rampant,§ Etienne Paux,*
Jacques Le Gouis,* Catherine Bastien,† and Vincent Segura†,1
*GDEC, INRA, UCA, 63000 Clermont-Ferrand, France, †BioForA, INRA, ONF, 45075 Orléans, France, ‡GenoBois analytical platform, INRA, 45075 Orléans, France, and §EPGV, INRA, CEA-IG/CNG, 91057 Evry, France

ORCID IDs: 0000-0003-0885-0969 (R.R.); 0000-0002-6029-0498 (J.-P.C.); 0000-0002-3094-7129 (E.P.); 0000-0001-5726-49
000-0002-9391-6637 (C.B.); 0000-0003-1860-2256 (V.S.)

Part III: Principle of phenomic selection, a low-cost and efficient alternative to genomic selection (Rincent, Segura et al., 2018)



Heffner *et al.* (2009)

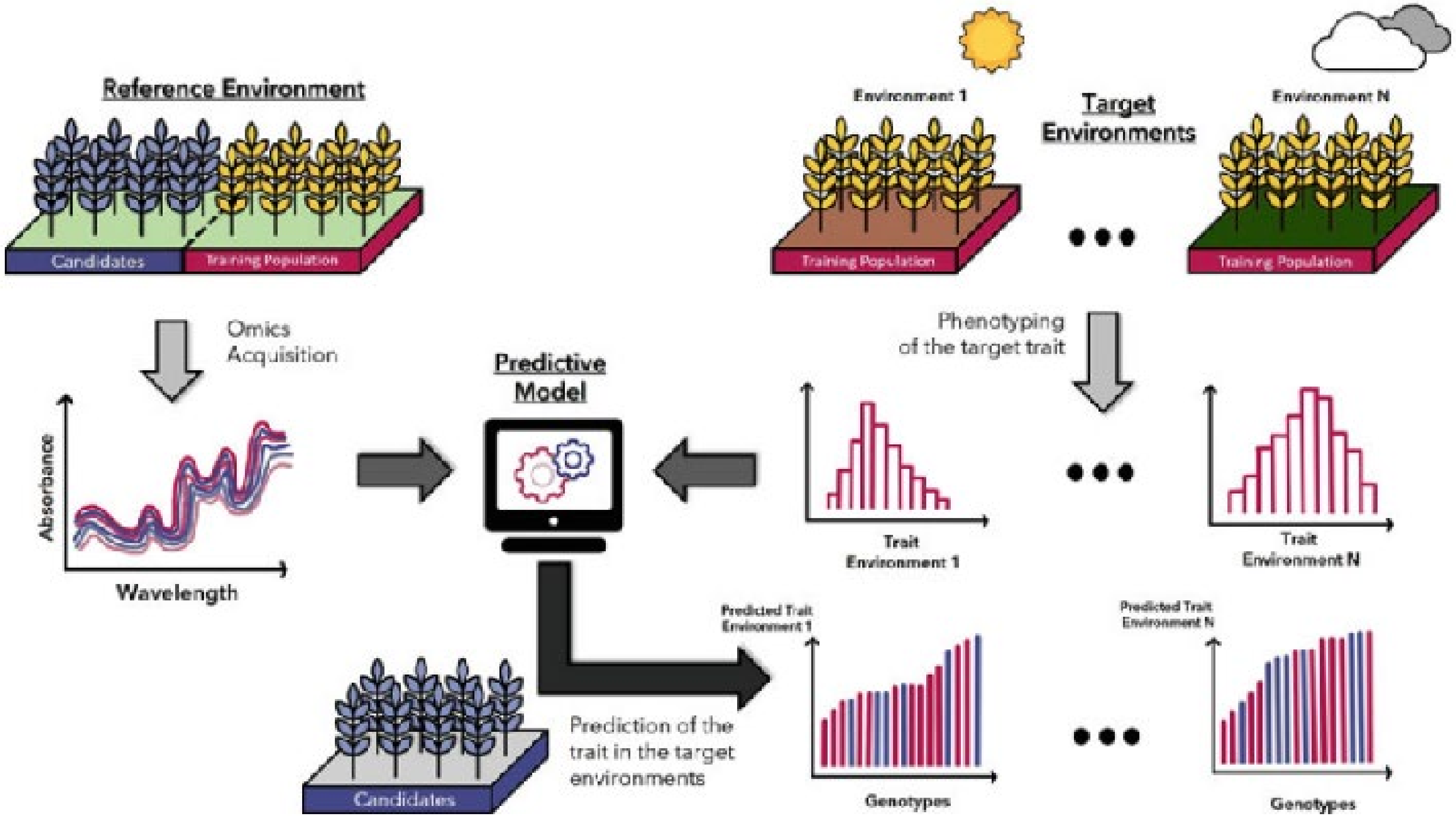


Estimation of a similarity matrix based on NIRs spectra instead of genomic data to calibrate model and predict trait using HBLUP

$$H(i,j) = \frac{\sum_{k=1}^{n_w} [S(i,k) \times S(j,k)]}{n_w},$$

Spectra are heritable, are under polygenic determinisms capture a part of GxE and

Part III: Principle of phenomic selection, a low-cost and efficient alternative to genomic selection (Rincent, Segura et al., 2018)



Robert et al., 2022

Estimation of a similarity matrix based on NIRs spectra instead of genomic data to calibrate model and predict trait using HBLUP

$$H(i,j) = \frac{\sum_{k=1}^{n_w}[S(i,k) \times S(j,k)]}{n_w},$$

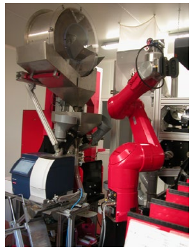
In Phenomic selection, NIR spectra **are not used to predict the chemical composition** of the sample (classical use of NIRS), but to **infer genetic covariance** between genotypes. Unlike with classical use of NIRS, in PS spectra are considered **at the genotypic scale**.

POC of using phenomic selection based on NIRs of landraces stored genebank to predict trait of landraces in contrasted environments - MineLandDiv and MALANIRs

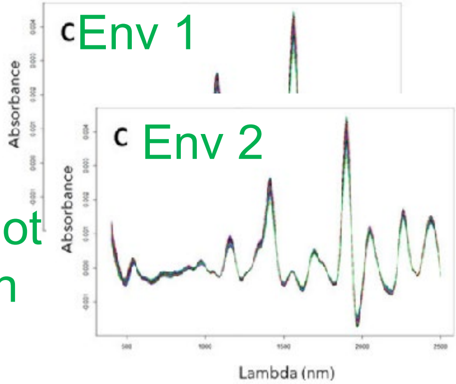
MineLandDiv
Mining allelic diversity of maize landraces for tolerance to abiotic and biotic stresses

Traits (incuding biochemical composition of grain)

UE Diascope



NIRs from grain harvested in each plot for 300 landraces in different trials with automated devices



Estimation of similarity between NIR spectra

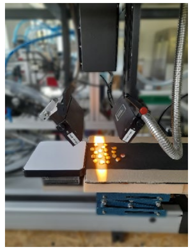
$$H(i,j) = \frac{\sum_{k=1}^{n_w}[S(i,k) \times S(j,k)]}{n_w},$$

Train a HBLUP model using phenotyping data from trial to predict trait

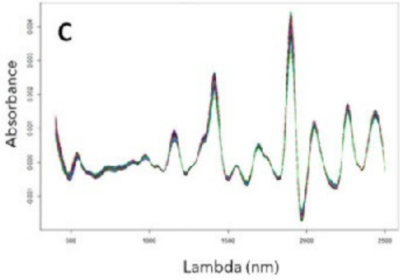
$$y = Xb + Za + \varepsilon$$

Genebanks
(ECPGR MALANIRs,
AGAP Institute)

NIRs of new landraces maintained in genebank



NIRs of seeds from genebank



Estimation of NIR spectra similarity of grain with NIR spectra from field trial

Predict trait (including biochemical composion of grain) of landraces maintained in genebank based on their NIR spectra

Low cost characterization of landraces in contrasted environment

PRO-GRACE/EMPHASIS SYMPOSIUM
27-28 June 2024

CONCLUDING REMARKS

- ✓ Genomic selection allowed to predict traits of landraces from genebanks based on allelic frequencies estimated in DNA bulk genotyped with 50K array
  - ➢ A way to characterize agronomic trait of landraces in genebank based on their DANS bulk genotyping with 50K

- ✓ NIRs is a low cost non destructive approach are currently used for predicting biochemical composition of seeds/silage harvested in field in the framework of maize breeding program
  - ➢ GWAS on seed composition based on NIRs equation identified QTL pointed to know genes as Ln1 for linoleic content
  - ➢ Genomic selection model trained on seed composition based on NIRs equation allowed to predict with good accuracy seed composition of landraces based on their genotyping with 50K
  - ➢ Combining NIRs with genomic data is a way to characterize seed composition of landraces maintained in genebank but also genetic determinism of trait.

- ✓ Phenomic selection based on NIRs could be a way to predict agronomic trait of landraces in genebank based on NIRs of seed maintained in genebank (MineLandDiv and MALANIRs project)
  - ➢ Automated devices for high troughput phenotyping of seeds by NIRs

Acknowledgements

Génétique Quantitative et Évolution Le Moulon

Charcosset A.
Moreau L.
Madur D.
Bauland C.
Combes V.
Ben Sadoun S.
Arca M.
Mary-Huard T.
Roux A.
Fievet J.
Galaretto A.O
Rincent R.
De Oliviera. Y
M. Tenaillon

agap

Gouesnard B.
Zanetto A.
Teres P.
Vincent M.
Diaw Y.
Ecarnot M.

EPGV

Faivre-Rampant P.
Hissinger D.
Bérard A.
Chevreau A.

UE SMH
Palaffre C.

UE Diascope
R. Chapuis
F. Meunier
P. Sartre

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

100 CENTENARIO MBG
MISIÓN BIOLÓGICA DE GALICIA

A. Butron
P. Revilla
R.A Malvar

crea
Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria

C. Balconi
A. Torri
R. Redaelli
G. Mazellini

ECPGR
S. Goritschning

Funding

- Amaizing
- Diversity Zea
- ECPGR EVA maize
- Dromamed
- MineLandDiv

amaizing

INVESTISSEMENTS D'AVENIR

European Cooperative Programme for Plant Genetic Resources
ECP/GR

INRAE

MineLandDiv
Mining allelic diversity of maize landraces for tolerance to abiotic and biotic stresses

FranceAgriMer

AGENCE NATIONALE DE LA RECHERCHE ANR

PROmaïs

AND all people that collected, maintained landraces and inbred lines collection in different European Genebanks and in France Center of genetic resources CRB Gamèt at Montpellier and lines collection at Experimental unit Saint Martin de Hinx

Génétique Quantitative et Évolution Le Moulon

IDEEV
Institut Diversité, Ecologie et Evolution du Vivant

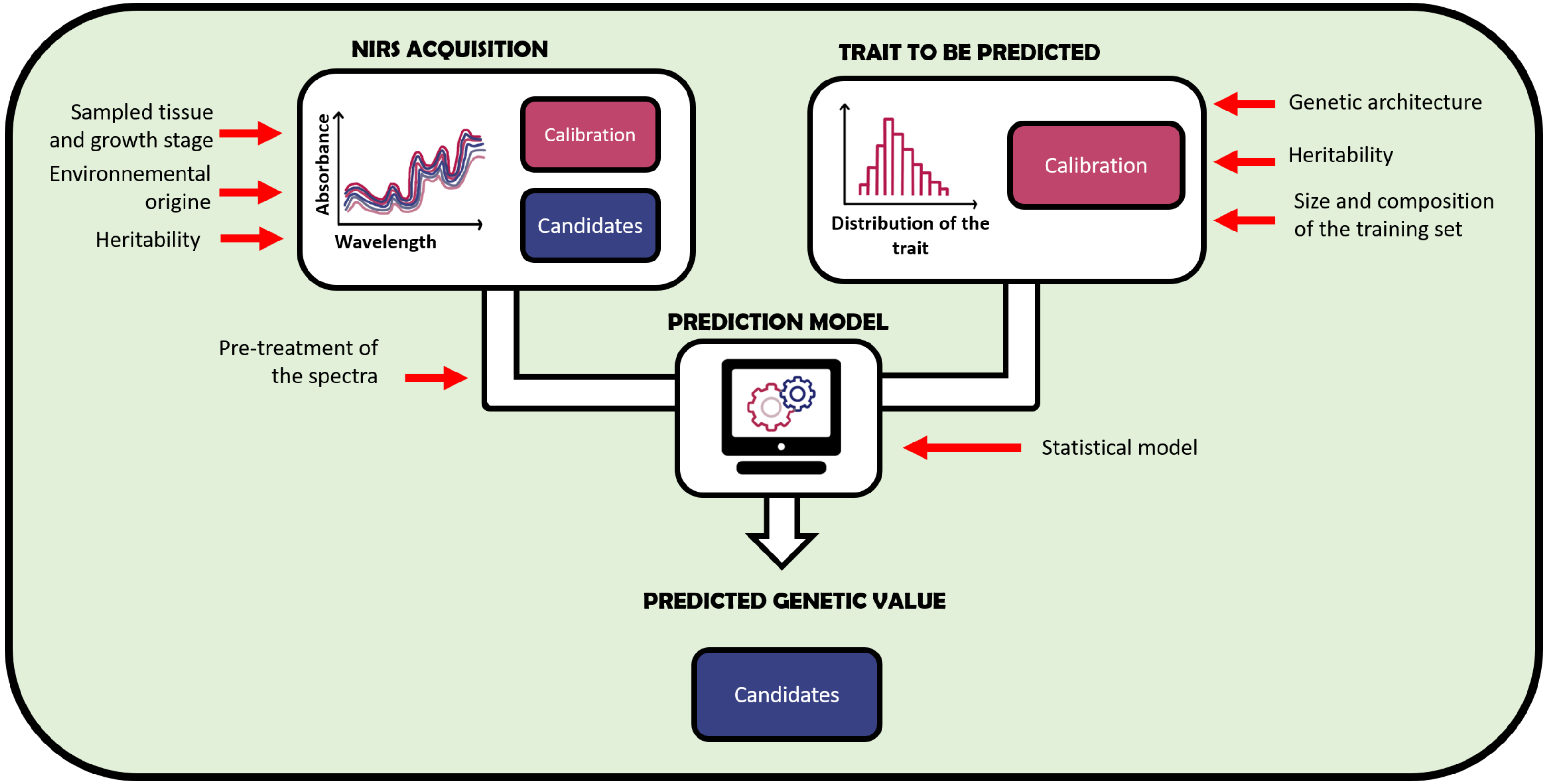
université PARIS-SACLAY FACULTÉ DES SCIENCES D'ORSAY

INRAE

cnrs

AgroParisTech
Talents d'une planète soutenable

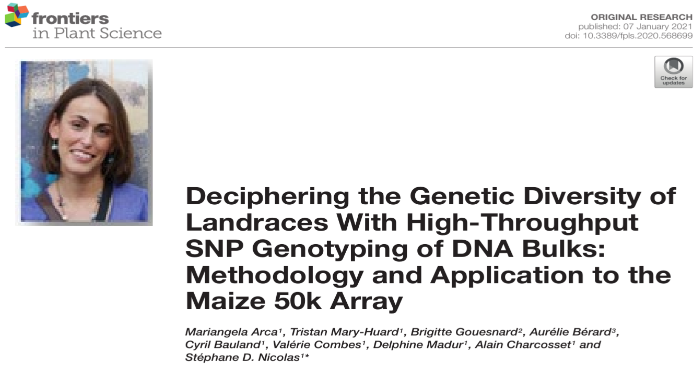
Part III: Principle of phenomic selection, a low-cost and efficient alternative to genomic selection (Rincent, Segura et al., 2018)



Adapted from P. Robert

In Phenomic selection, NIR spectra **are not used to predict the chemical composition** of the sample (classical use of NIRS), but to **infer genetic covariance** between genotypes. Unlike with classical use of NIRS, in PS spectra are considered **at the genotypic scale**.

INTRODUCTION: Development of a new pooling method based on 50K array to genotype landraces: A two-step approach



frontiers in Plant Science

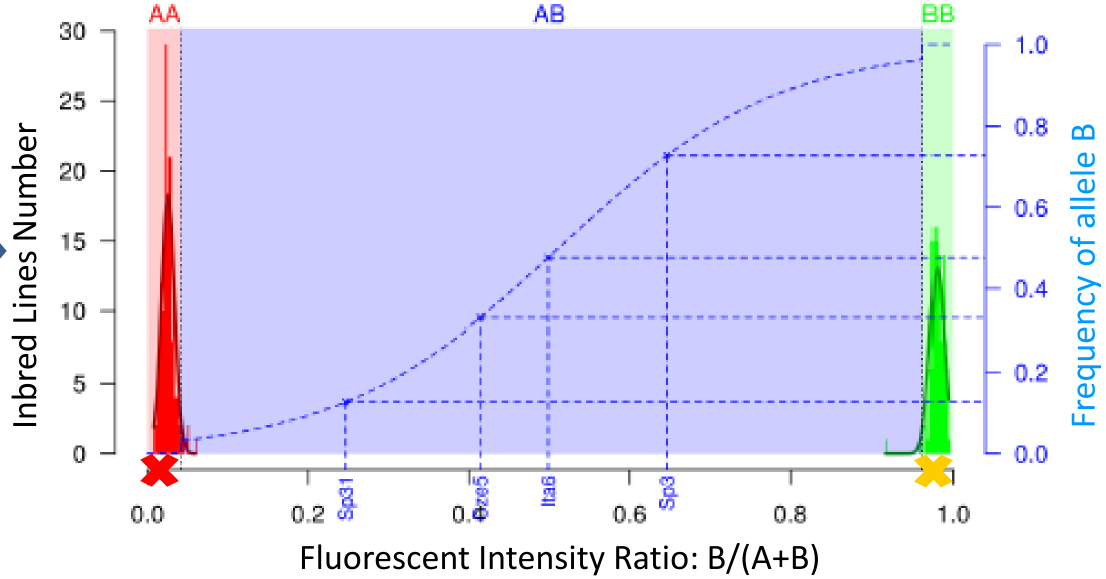
ORIGINAL RESEARCH
published: 07 January 2021
doi: 10.3389/fpls.2020.568699

Deciphering the Genetic Diversity of Landraces With High-Throughput SNP Genotyping of DNA Bulks: Methodology and Application to the Maize 50k Array

*Mariangela Arca¹, Tristan Mary-Huard¹, Brigitte Gouesnard², Aurélie Bérard³, Cyril Bauland¹, Valérie Combes¹, Delphine Madur¹, Alain Charcosset¹ and Stéphane D. Nicolas¹**

Step 1: Test of landraces homozygosity for A and B alleles based on the distributions of FIR from AA or BB inbred lines

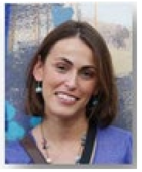
Step 2: Predict allelic frequencies using a unique logistic model for all SNP trained on 2 controlled pools with known allelic frequencies and 1000 SNPs

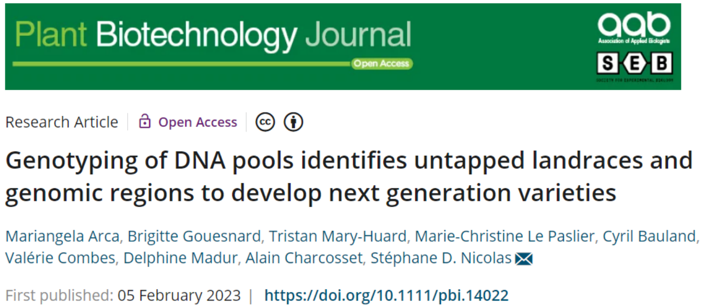


	Frequency of allele B		
Landraces	SNP1	SNP2	SNP3
Landrace X	20%	30%	50%
Landrace Y	70%	0%	30%

Accurate estimation of allelic frequencies within landraces with mean absolute error of 3% / Fluorescent intensity ratio is highly reproducible between labs ($r^2 = 0.87$)

INTRODUCTION: Spatial genetic structuration of genetic diversity at worldwide level (Arca et al., 2022)





Plant Biotechnology Journal
Open Access

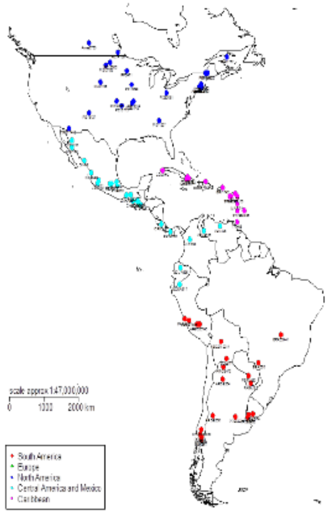
Research Article | Open Access

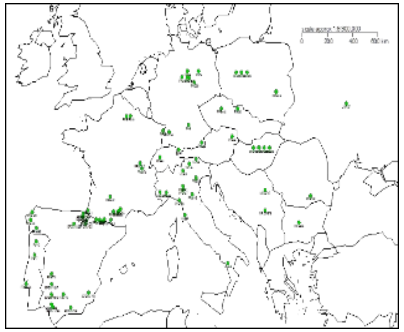
Genotyping of DNA pools identifies untapped landraces and genomic regions to develop next generation varieties

Mariangela Arca, Brigitte Gouesnard, Tristan Mary-Huard, Marie-Christine Le Paslier, Cyril Bauland, Valérie Combes, Delphine Madur, Alain Charcosset, Stéphane D. Nicolas

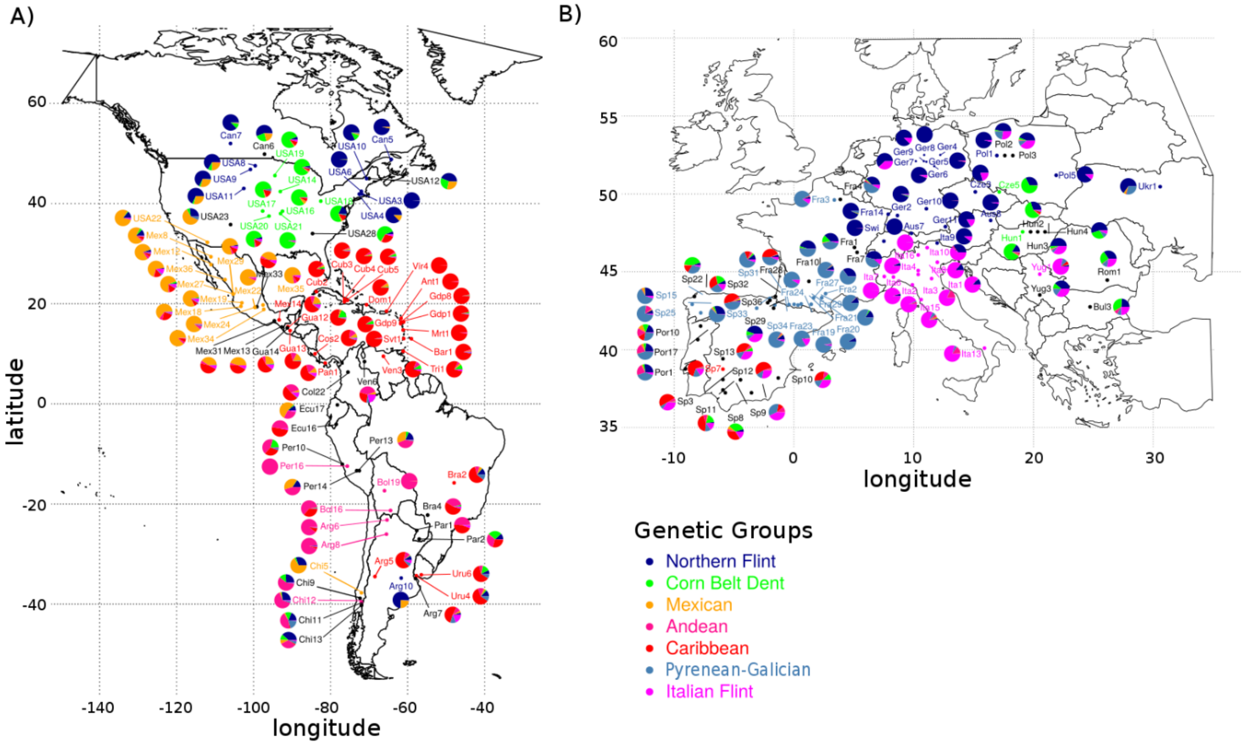
First published: 05 February 2023 | https://doi.org/10.1111/pbi.14022

156 landraces representing America and European diversity





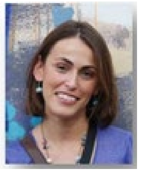
Structure analyses of 156 landraces using Admixture identified 7 genetic groups according to Evano Criterion:



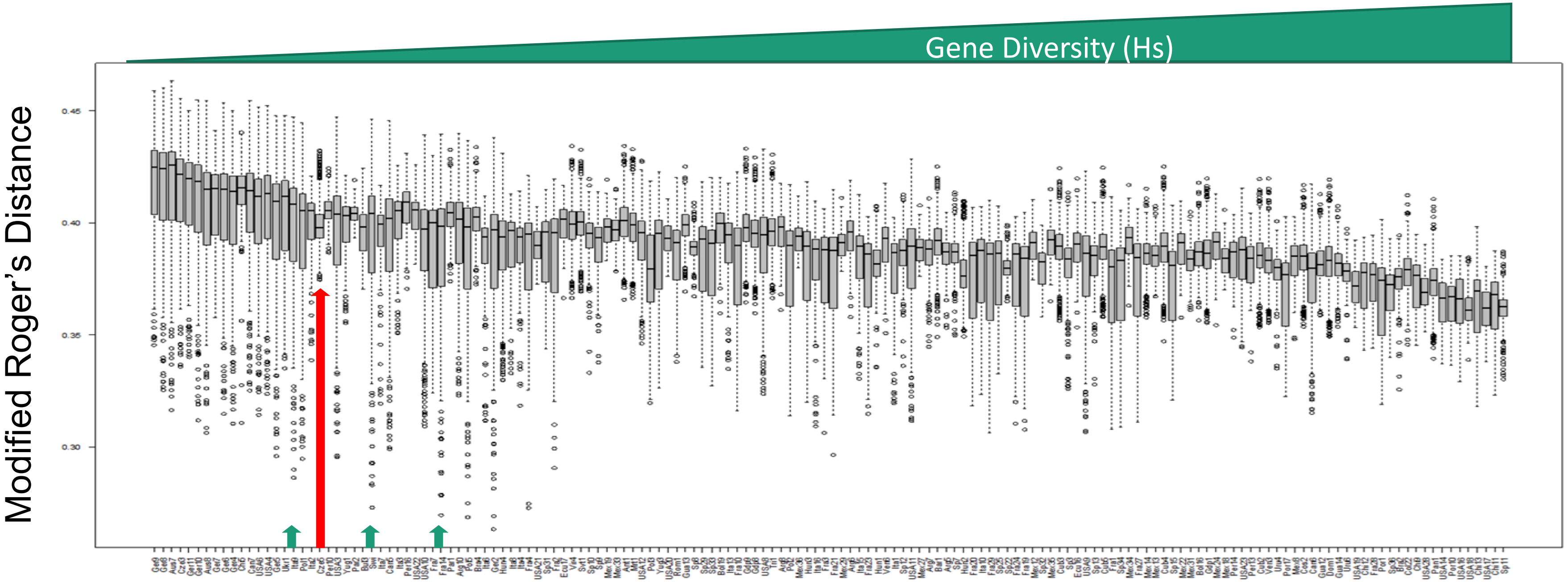
The genetic structuration highlighted clear geographical trends with latitude and longitude

=> Structuration were driven mostly by adaptation and isolation by distance (Highly consistent with previous report)

Relationship between 156 landraces and 327 CK lines (Arca et al., 2022)



Modified Roger’s Distance distribution between 327 CK lines and each landraces ordered according to their gene diversity (Nei, 1973)



While some landraces have a lot of closely related inbred lines,

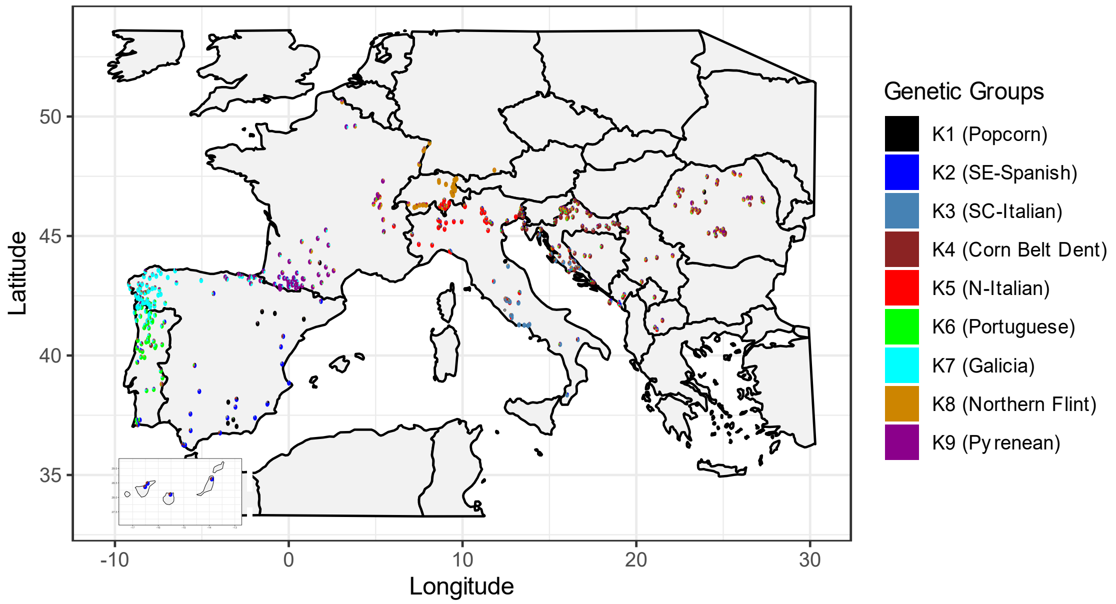
some other have not!

Some landraces poorly contributed to inbred line panel genetic diversity

INTRODUCTION: Extending 50K genotyping to european landraces to identify new sources of genetic diversity by comparing landraces from different genebanks (ECPGR EVA project)

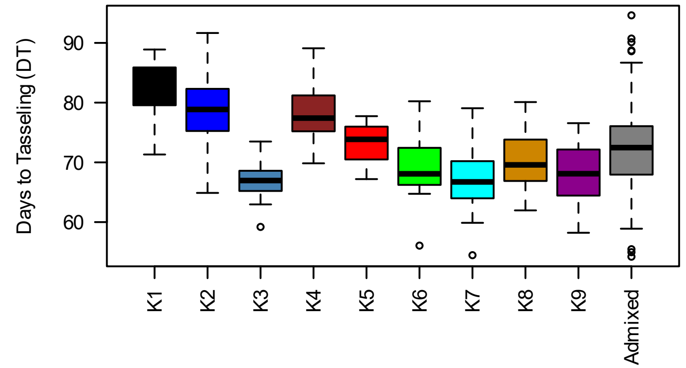
Carlotta, Galaretto et al., Biology, under review

Analysis of 623 European landraces from 9 European genebanks

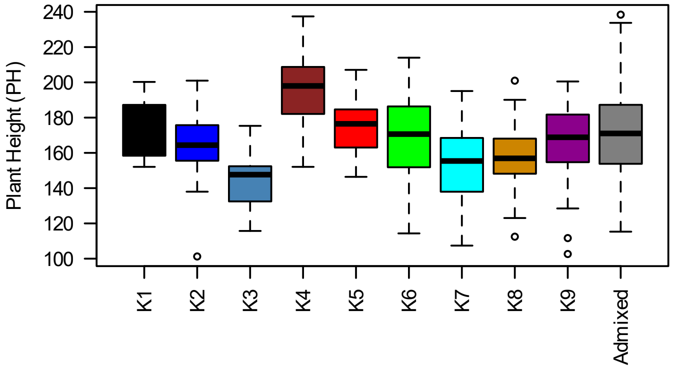


Evaluation of 623 European landraces for agronomic traits accross European Evaluation network (Common garden) for agronomic traits

DT by Genetic Group (K=9)



PH by Genetic Group (K=9)



1,526 landraces currently genotyped with 50K using DNA bulk => Database to identify promising sources of genetic diversity