# Genetic determinants of maize phenotypic evolution

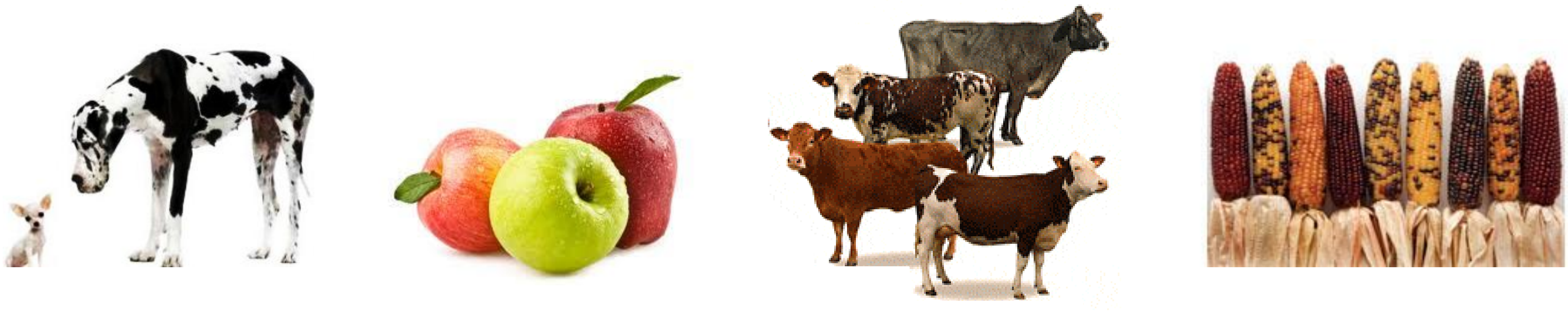
Le Moulon
GÉNÉTIQUE QUANTITATIVE ET ÉVOLUTION

IDEEV
Institut Diversité, Ecologie et Evolution du Vivant

Maud Tenaillon

cnrs

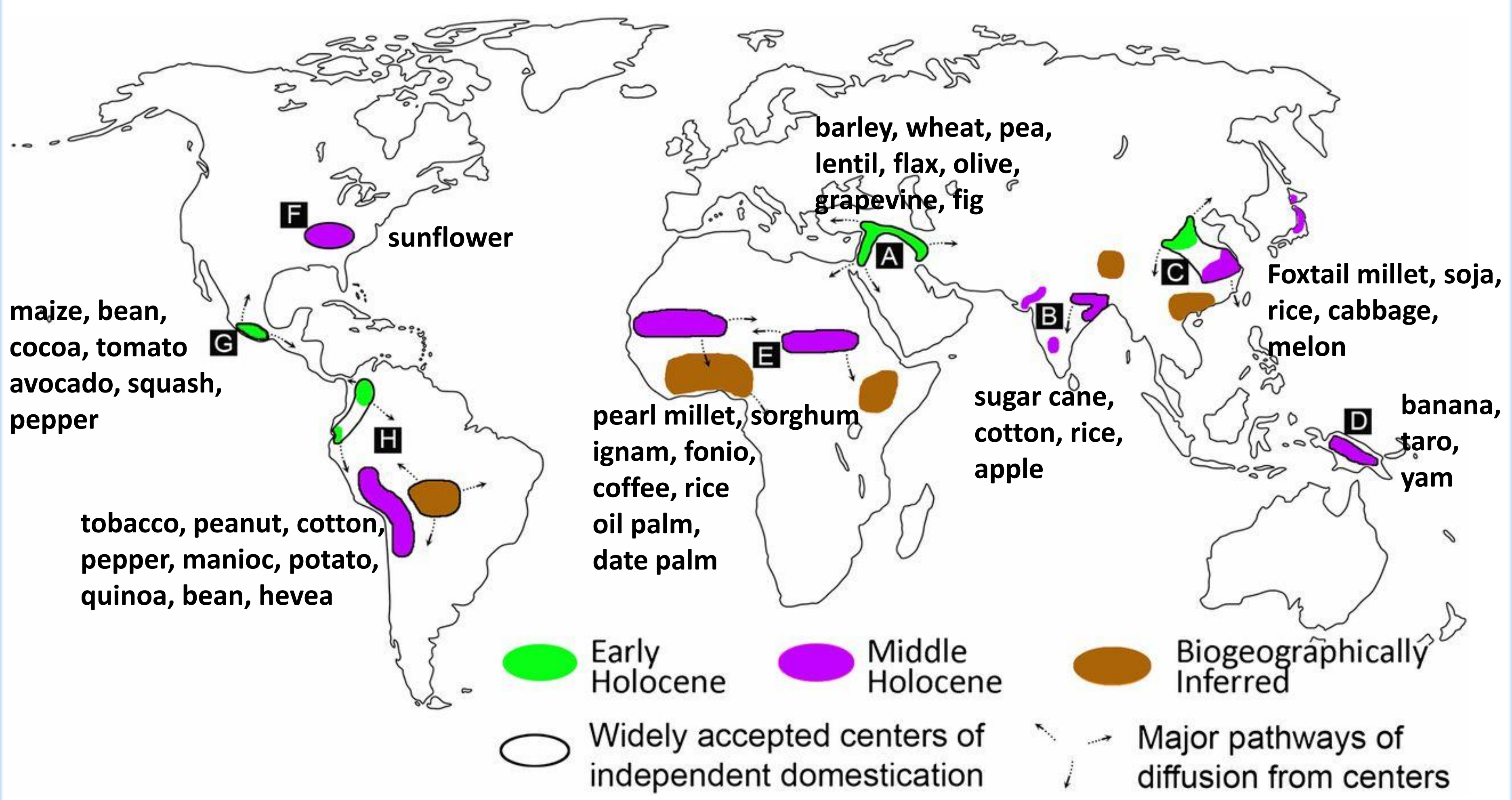
# Adaptation to human-needs in domesticated species



Domesticated species are good models to study adaptation because

- Long term *in vivo* evolution (>8,000 years)
- The variation is particularly pronounced
- Cultivated range of domesticated species by far exceeds that of their wild "ancestors"
- Genomic data for most domesticated species are produced

# Centers and epochs of plant domestication



barley, wheat, pea, lentil, flax, olive, grapevine, fig

sunflower

Foxtail millet, soja, rice, cabbage, melon

maize, bean, cocoa, tomato avocado, squash, pepper

sugar cane, cotton, rice, apple

pearl millet, sorghum ignam, fonio, coffee, rice oil palm, date palm

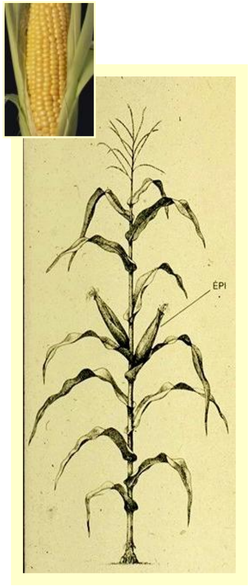
banana, taro, yam

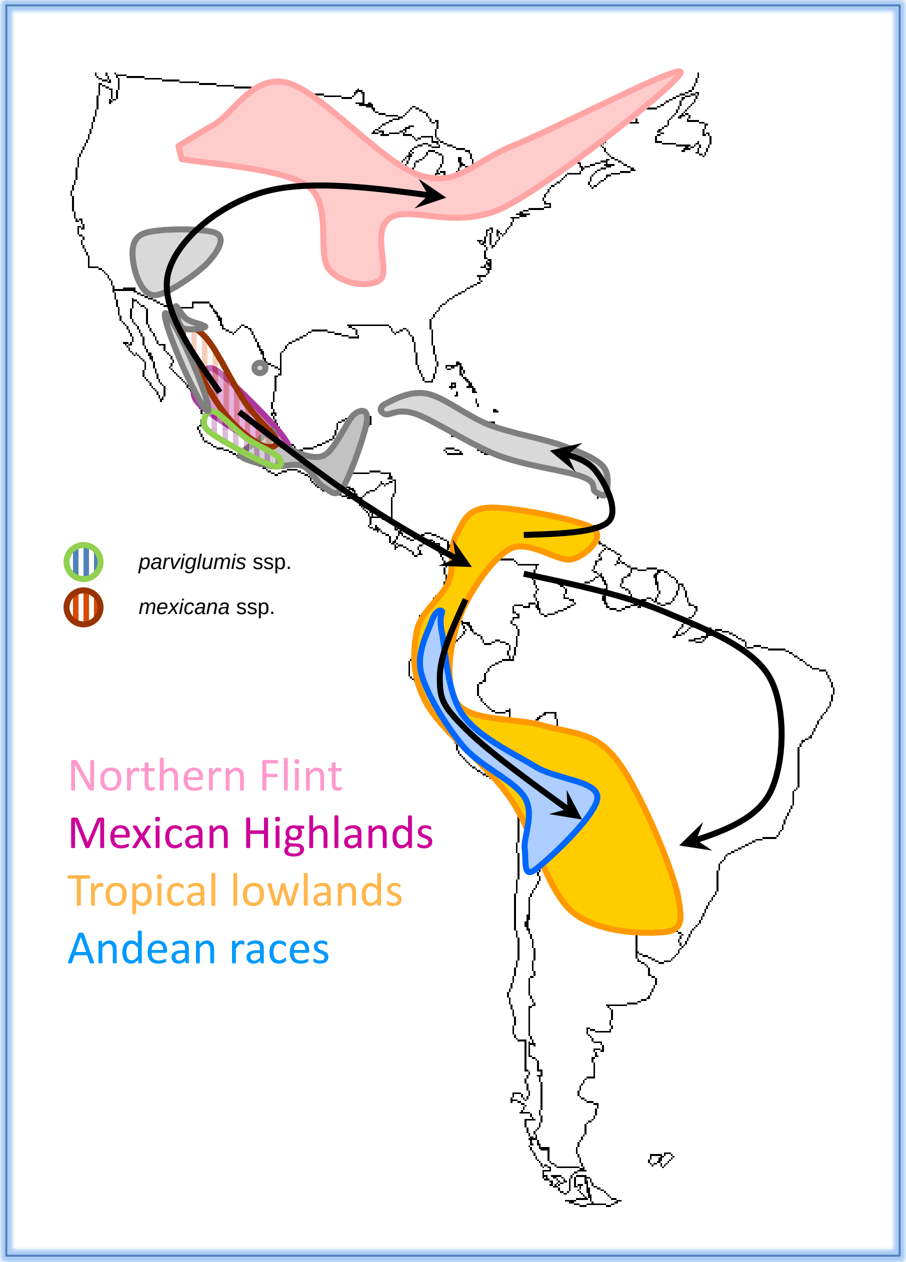
tobacco, peanut, cotton, pepper, manioc, potato, quinoa, bean, hevea

Early Holocene: -(12 000-8 200) years, Late Holocene: -(8 200-4 200) years

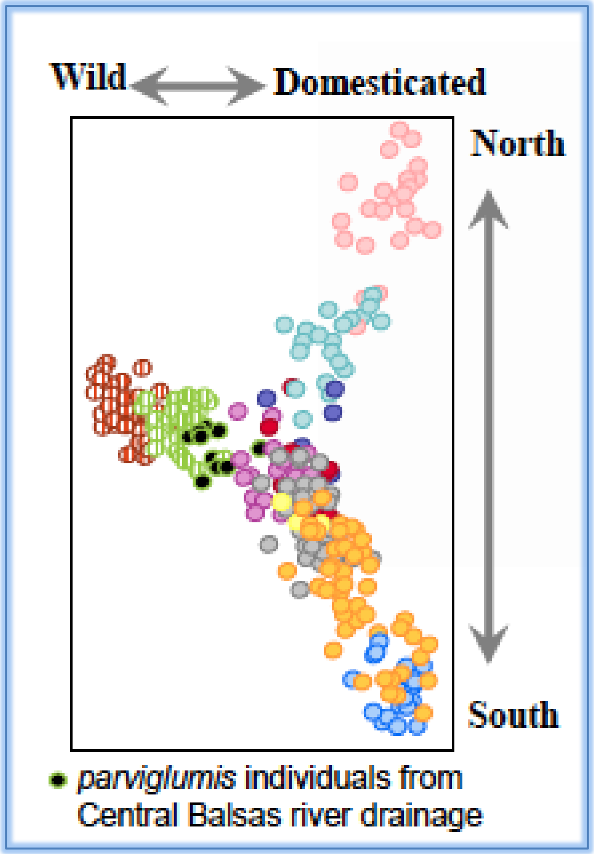
*Adapted from Larson et al. PNAS 2014*

# Origin and history of corn





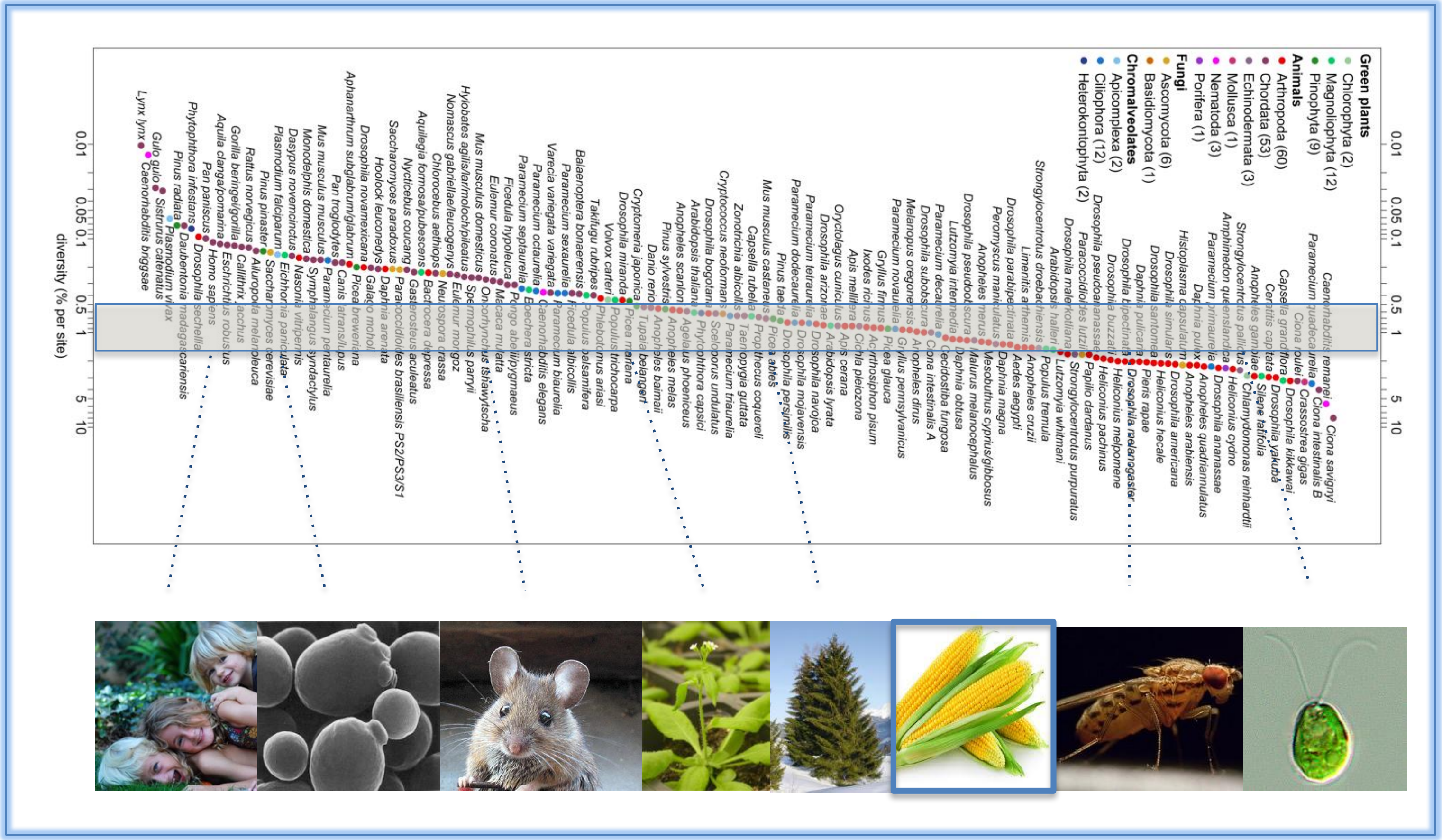
Domestication occurred in Mexico around 9000 years ago



● *parviglumis* individuals from Central Balsas river drainage

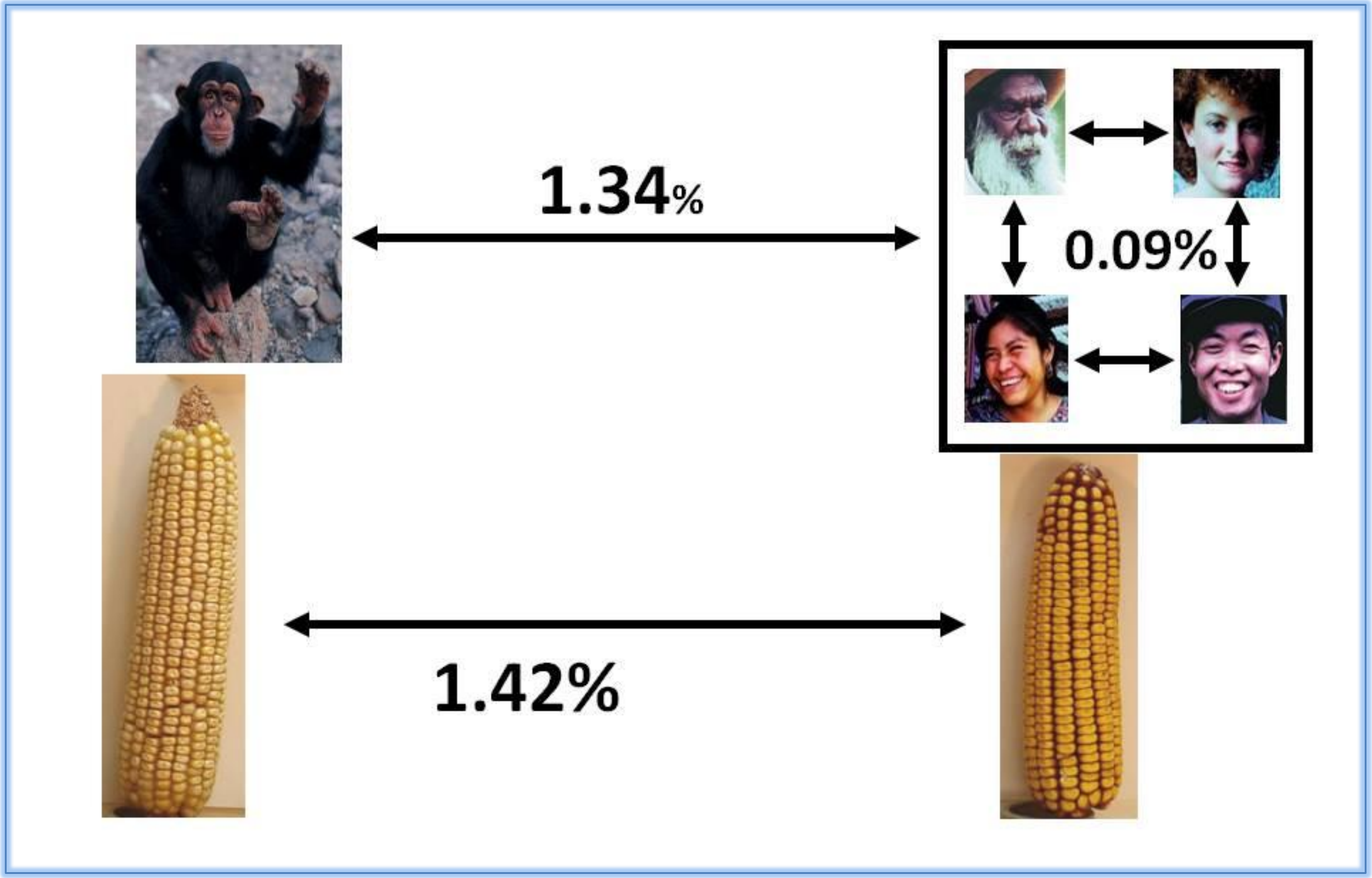
*Matsuoka et al. PNAS 2002*
*Tenaillon and Manicacci Adv. In Maize 2011*

# Genetic diversity in maize



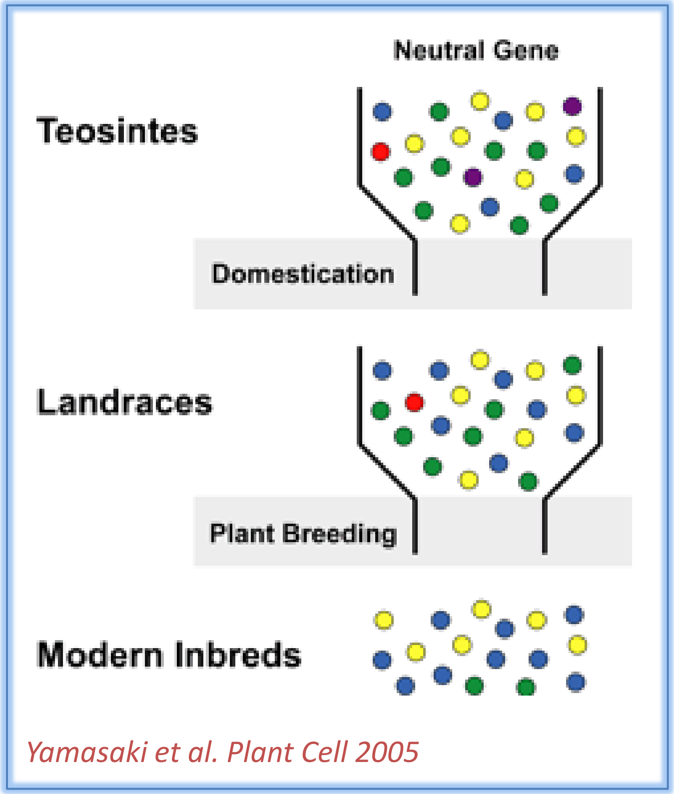
*Adapted from Leffler et al. Plos Bio 2012*

# Maize is a diverse model species

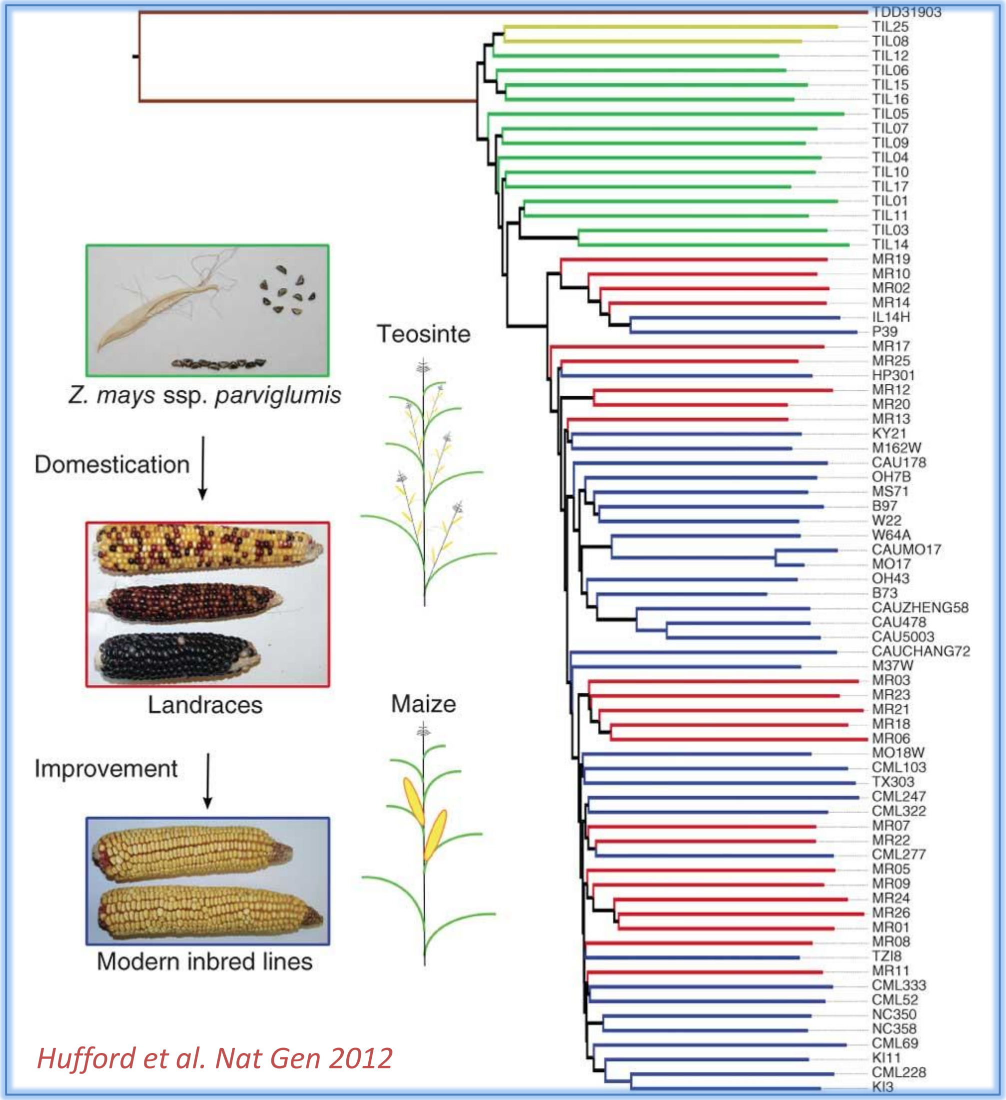


*Buckler et al. 2009*

# Maize underwent two bottlenecks



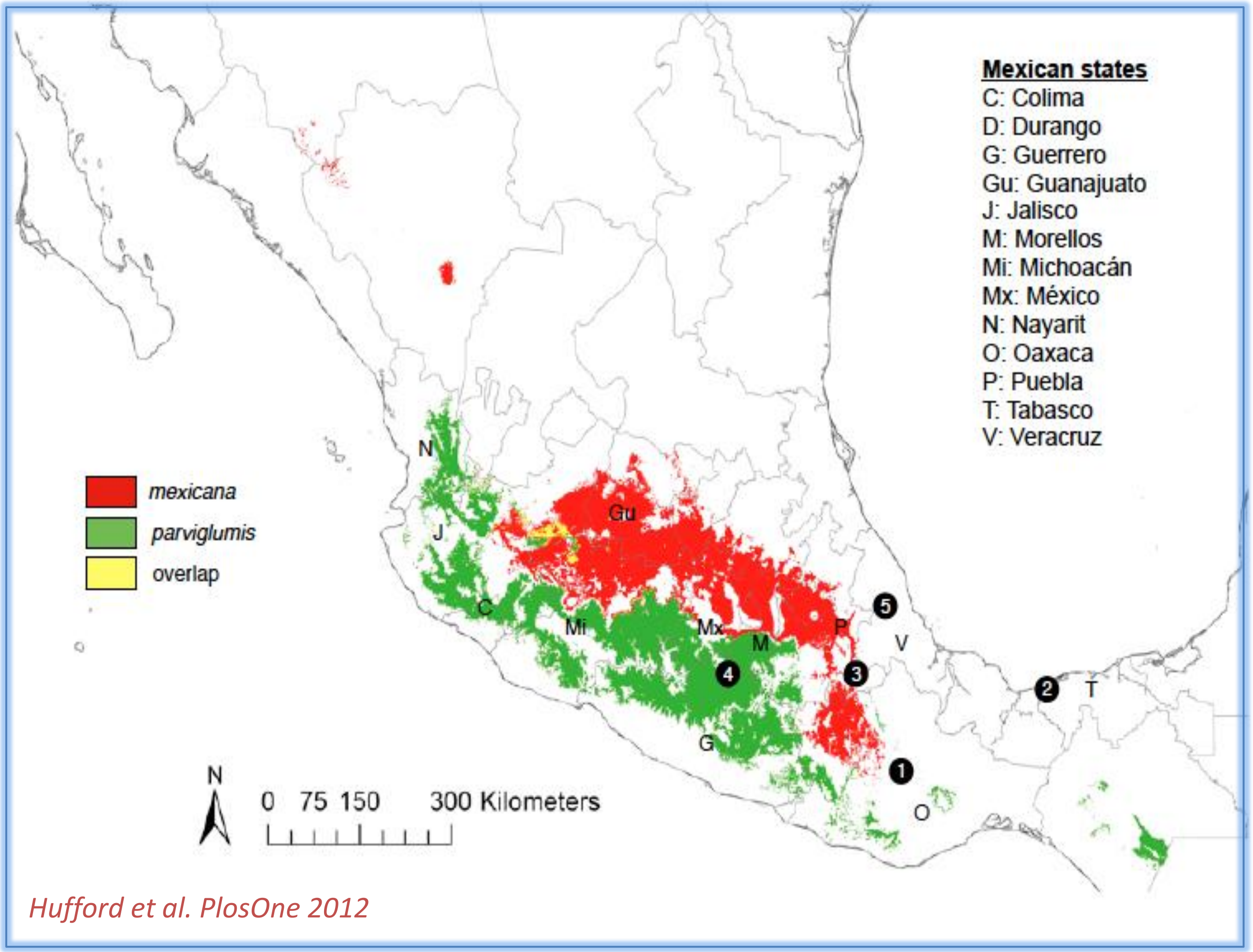
*Yamasaki et al. Plant Cell 2005*



*Hufford et al. Nat Gen 2012*

The reduction of diversity for the domestication bottleneck has been estimated to 20% and the breeding bottleneck to < 5%

# Niche modeling



*Hufford et al. PlosOne 2012*

Distinct ecological niches with little overlap

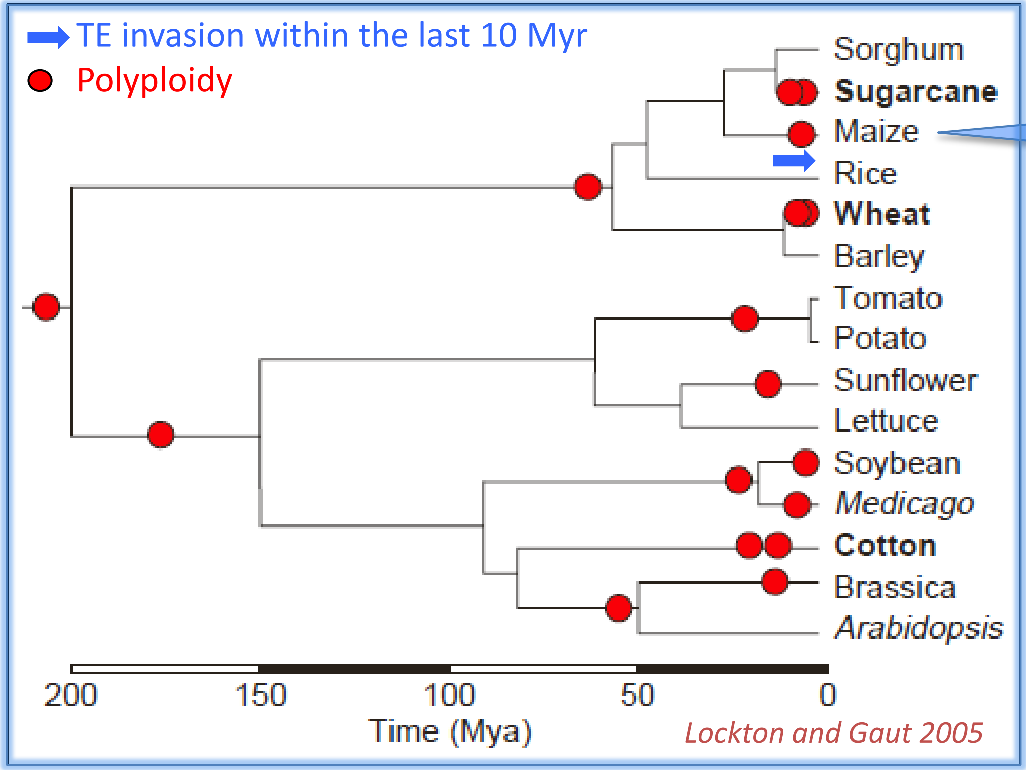
Average altitude for *parviglumis* 1058 m, for *mexicana* 2105 m

# DISTRIBUCION DE LOS PRINCIPALES MAICES CRIOLLOS

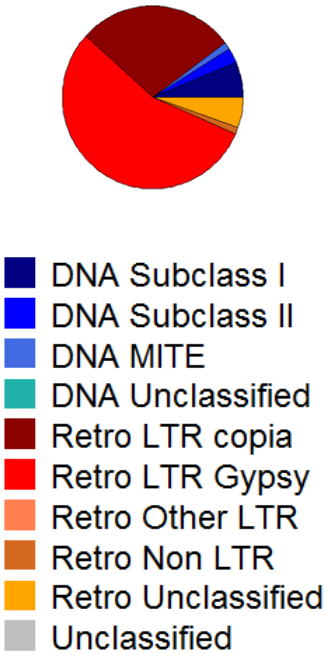
[illegible]

# The repeat fraction of the maize genome

## Maize genome is about 2.5 Gb, is highly redundant



*Lockton and Gaut 2005*



*Vitte et al. Brief Funct Gen 2014*

## and contains 85% of transposable elements

# Outline

- A bit more on the European history of maize

Drivers of maize adaptation:

- Short life cycle? Relation to genome size evolution
- Flowering time

# Outline

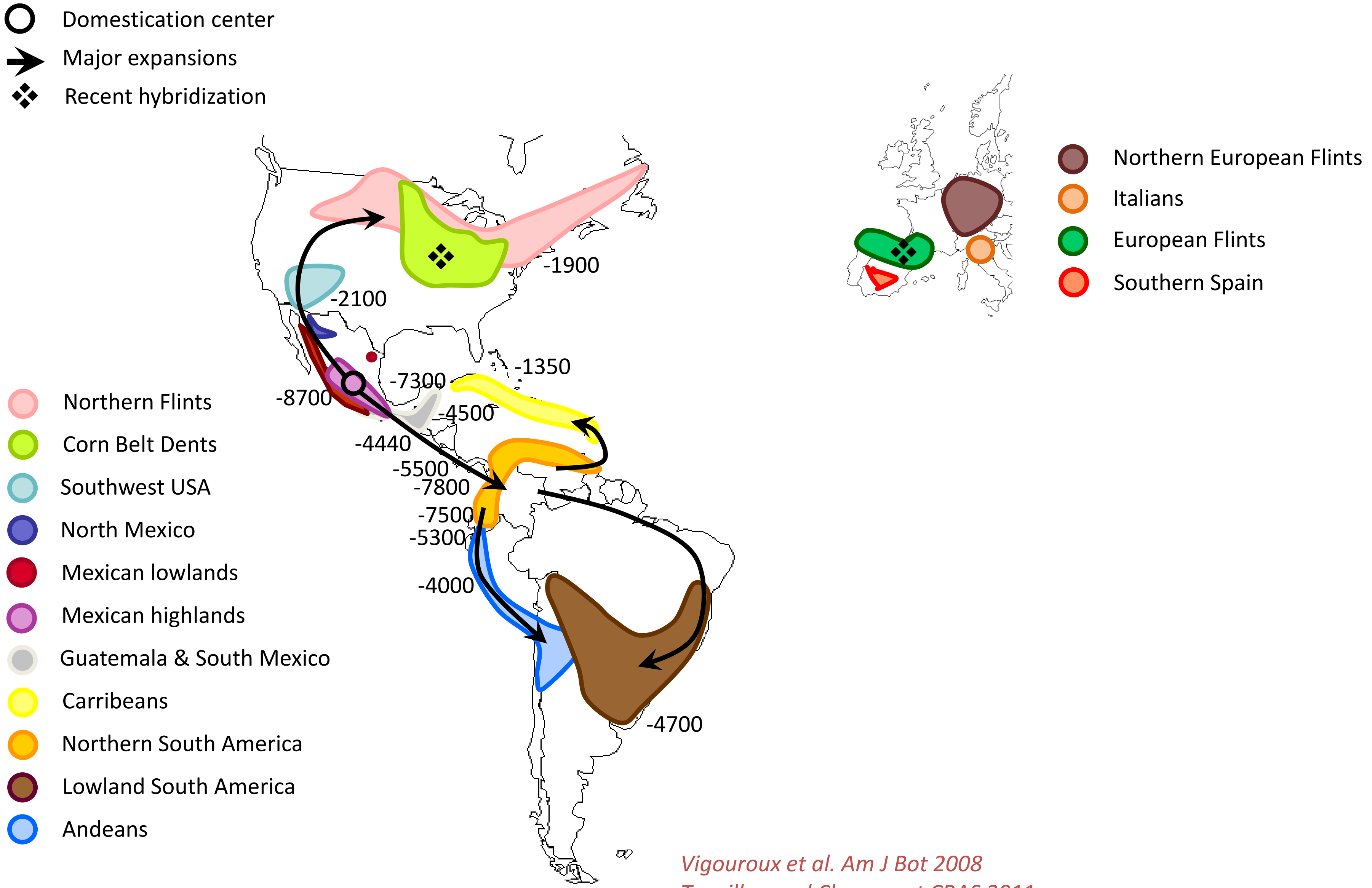
- A bit more on the European history of maize

J-T Brandenburg

Jean-Tristan Brandenburg, GQE-Le Moulon, France
Tristan Mary-Huard, GQE-Le Moulon, France
Guillem Riguaill, IPS2, France
Clémentine Vitte, GQE-Le Moulon, France
Johann Joets, GQE-Le Moulon, France
Stéphane Nicolas, GQE-Le Moulon, France
Alain Charcosset, GQE-Le Moulon, France

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR
INVESTISSEMENTS D'AVENIR

# Origin and history of corn



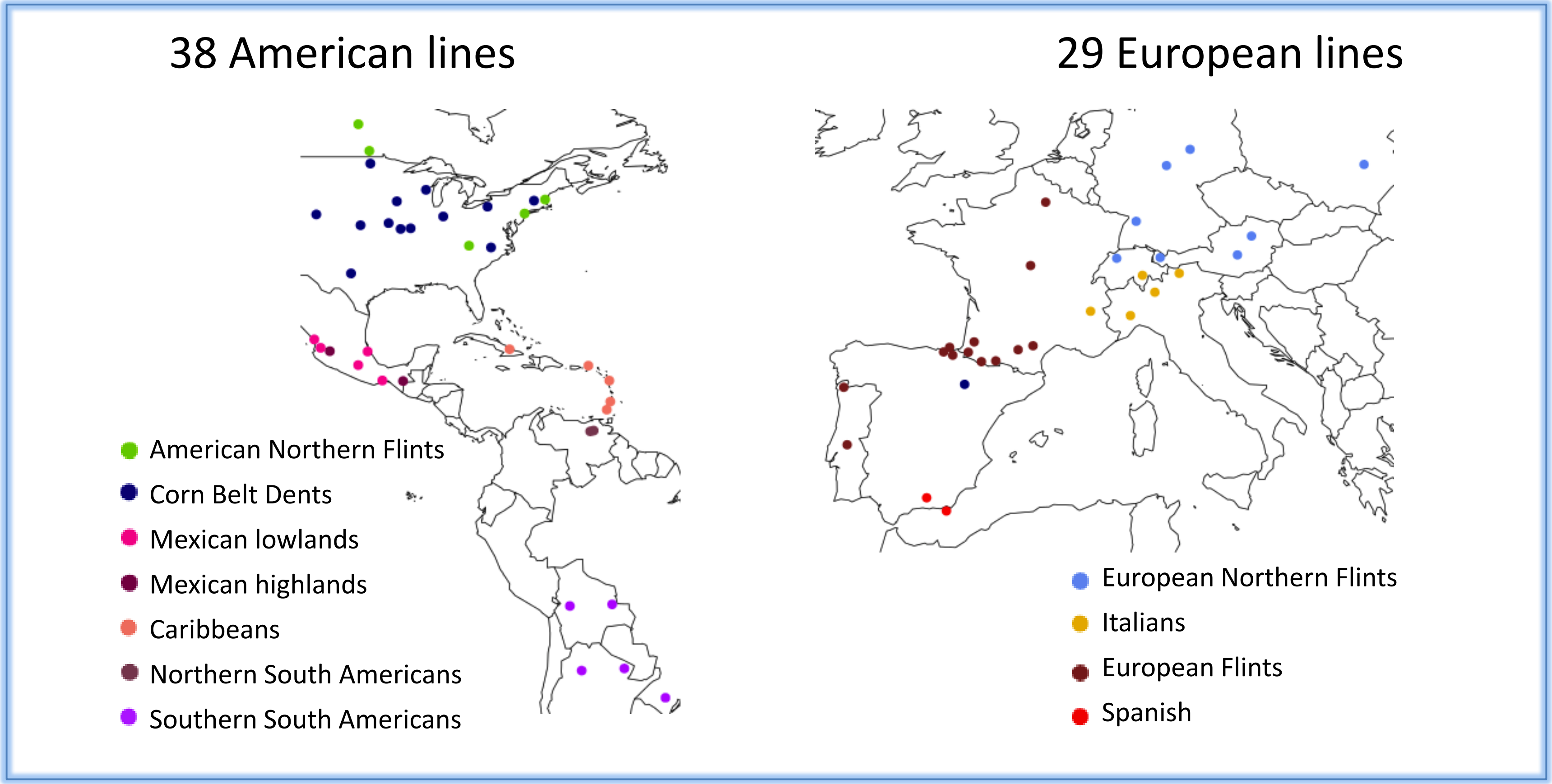
*Vigouroux et al. Am J Bot 2008*
*Tenaillon and Charcosset CRAS 2011*

# Sampling

Favour inbred lines directly derived from landraces

Ensure a correct representation of European genetic groups

Include all possible American sources of European maize



38 American lines

29 European lines

First-cycle inbreds + Single seed descents (57), Doubled haploids (10)

# Methodology

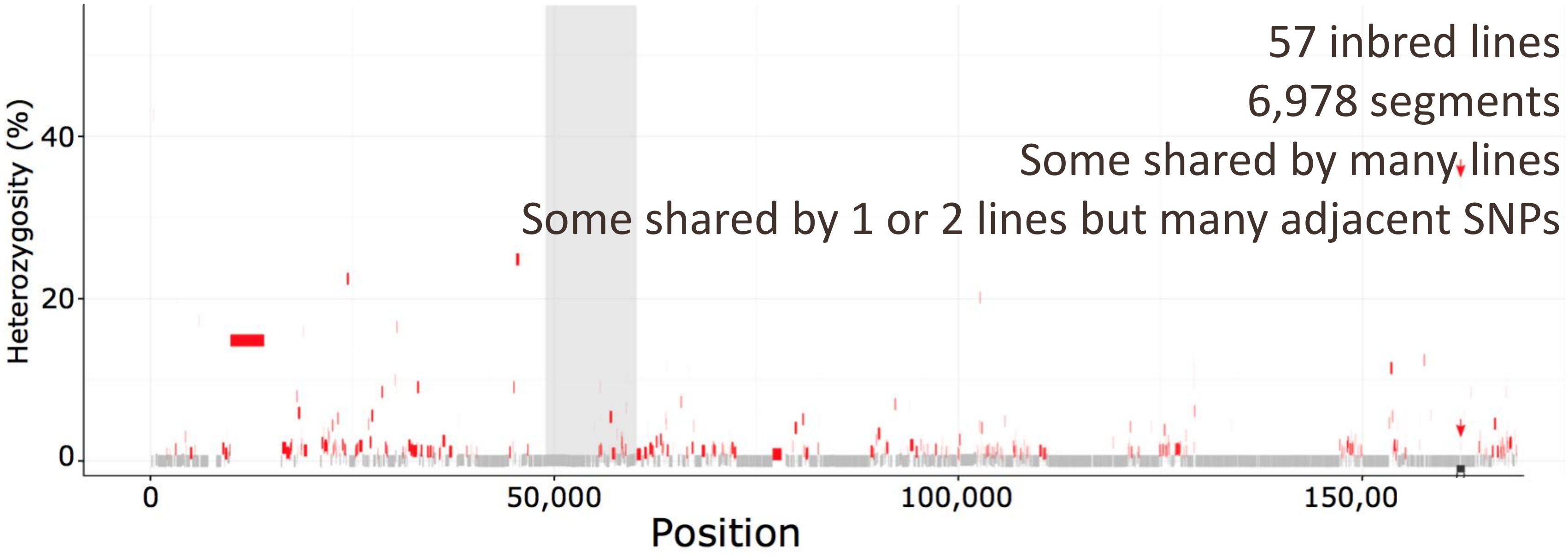
Sequencing:

67 mid-depth whole genome sequencing (18x)
>22 million SNPs genotyped
covering >85% of all maize genes
Alignment to an outgroup: *Tripsacum dactyloides*

False positive rate for homozygotes calls is 0.036%
False positive rate for heterozygotes calls is 34%

# Patterns of heterozygosity in inbreds

Detection of segments with unexpectedly high level of heterozygosity along chromosomes



57 inbred lines
6,978 segments
Some shared by many lines
Some shared by 1 or 2 lines but many adjacent SNPs

79% are unique to inbreds (not detected in Doubled Haploids)

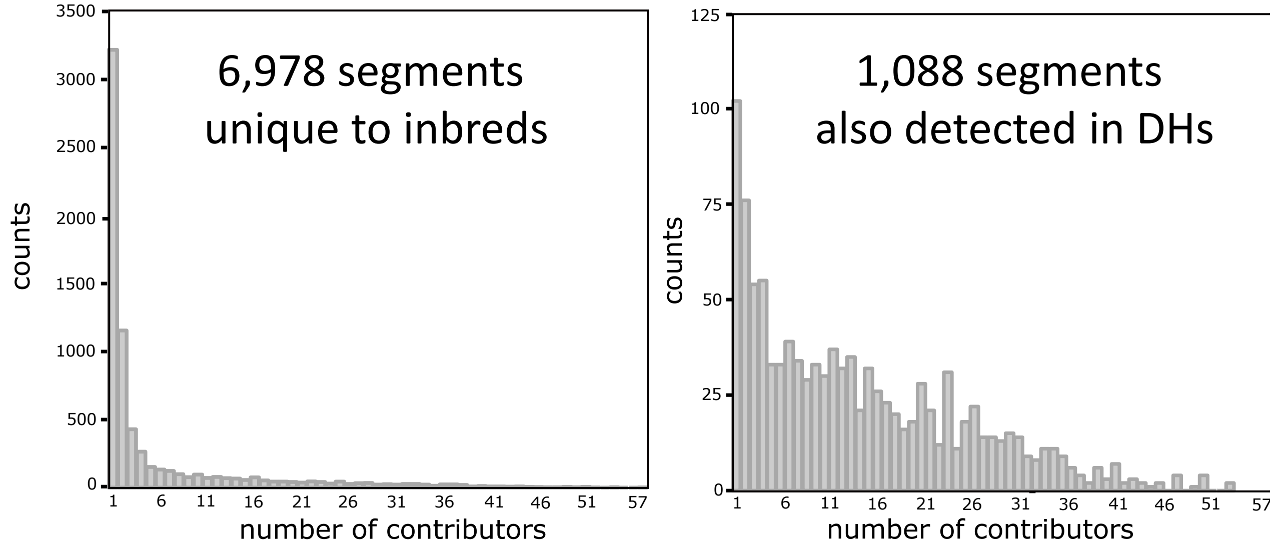
SIFT: heterozygous segments encompass significantly more deleterious variants than the rest of the genome

Selection against inbreeding depression contributes to shape patterns of residual heterozygosity

# Patterns of heterozygosity in Doubled Haploids

21% of segments (1,088) are found in DHs = alignment artefacts due to structural variation

These segments share a greater number of contributors

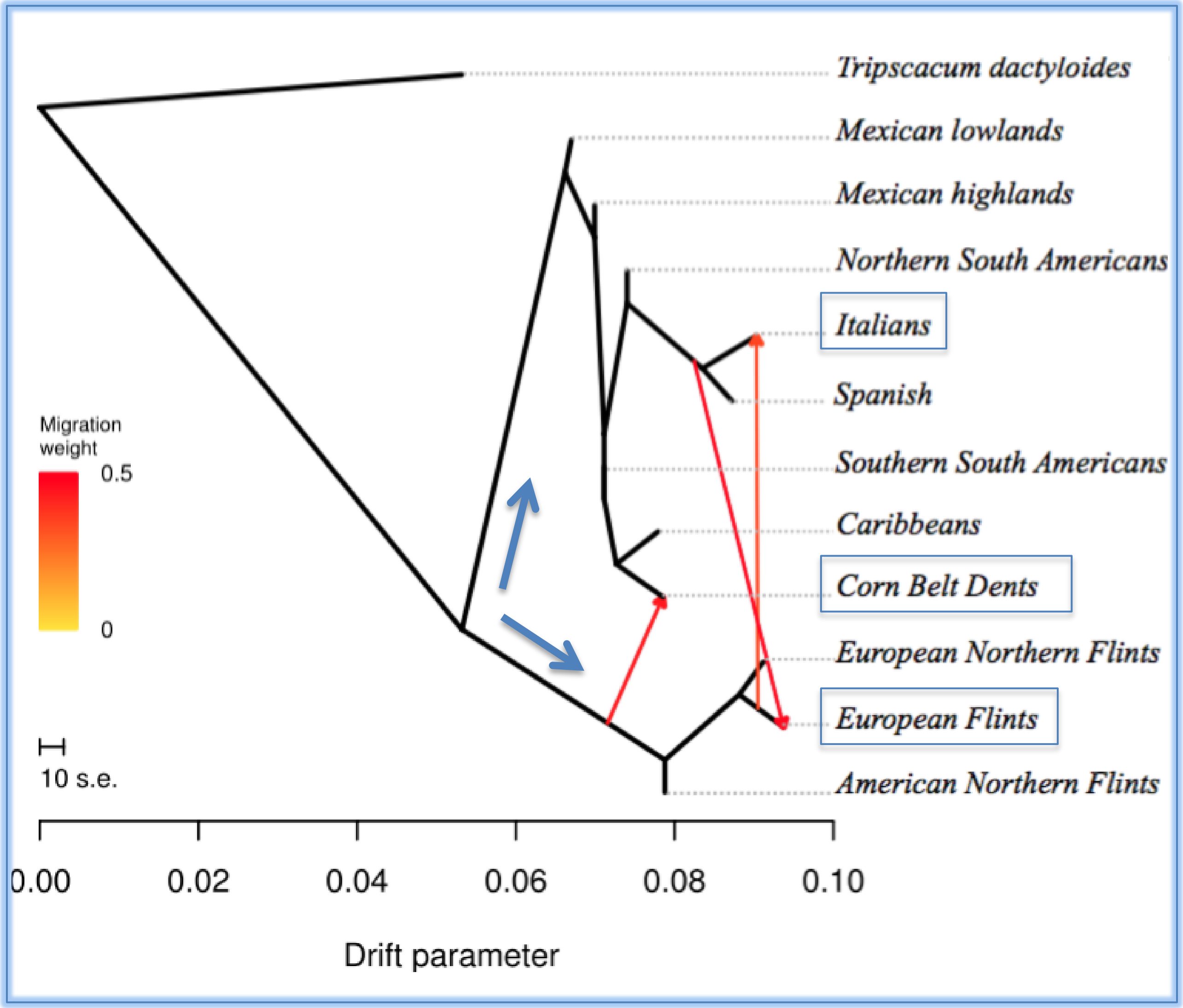


6,978 segments unique to inbreds

1,088 segments also detected in DHs

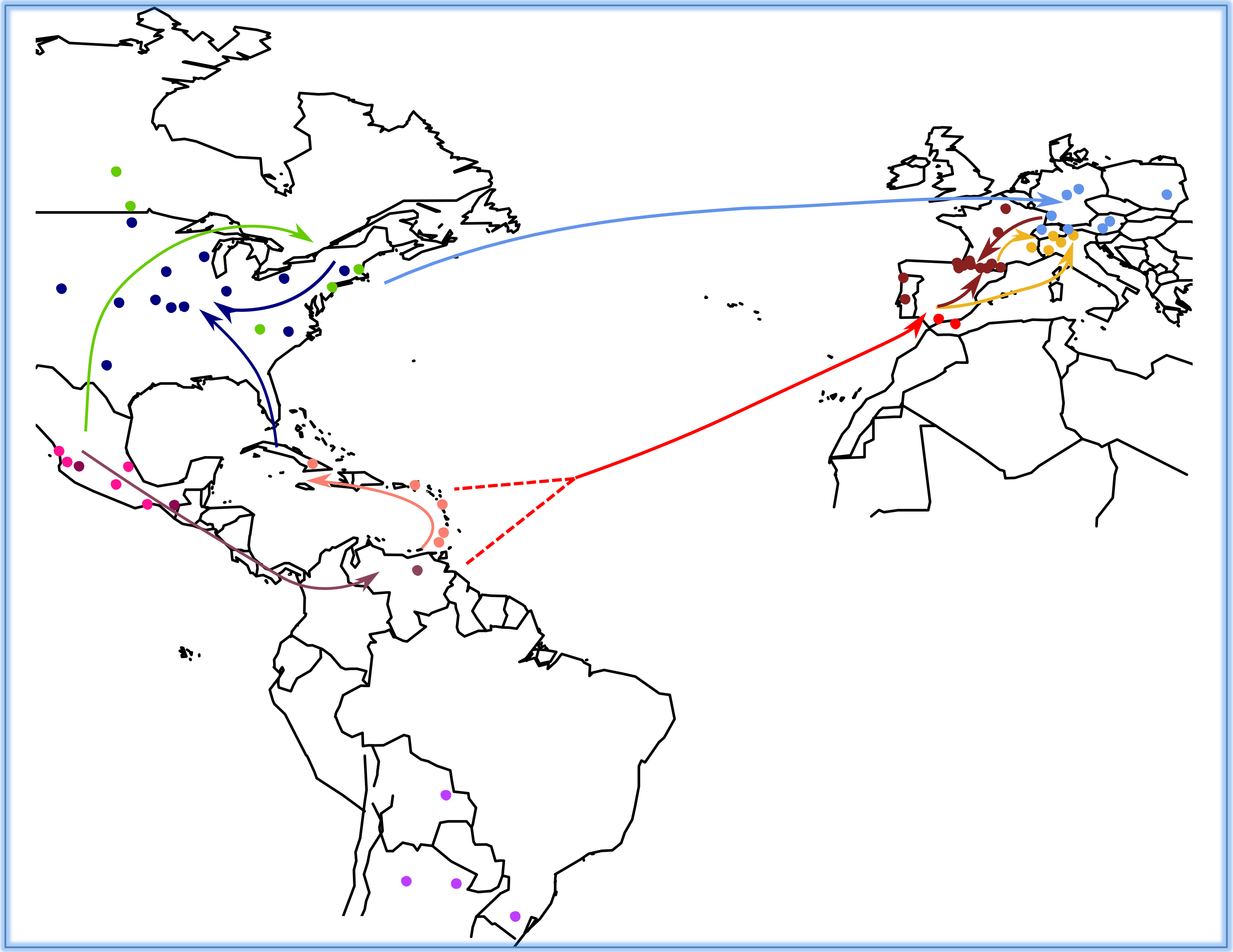
Substantial proportion of stretches of heterozygosity are caused by common structural variants

# Genetic proximities and admixture



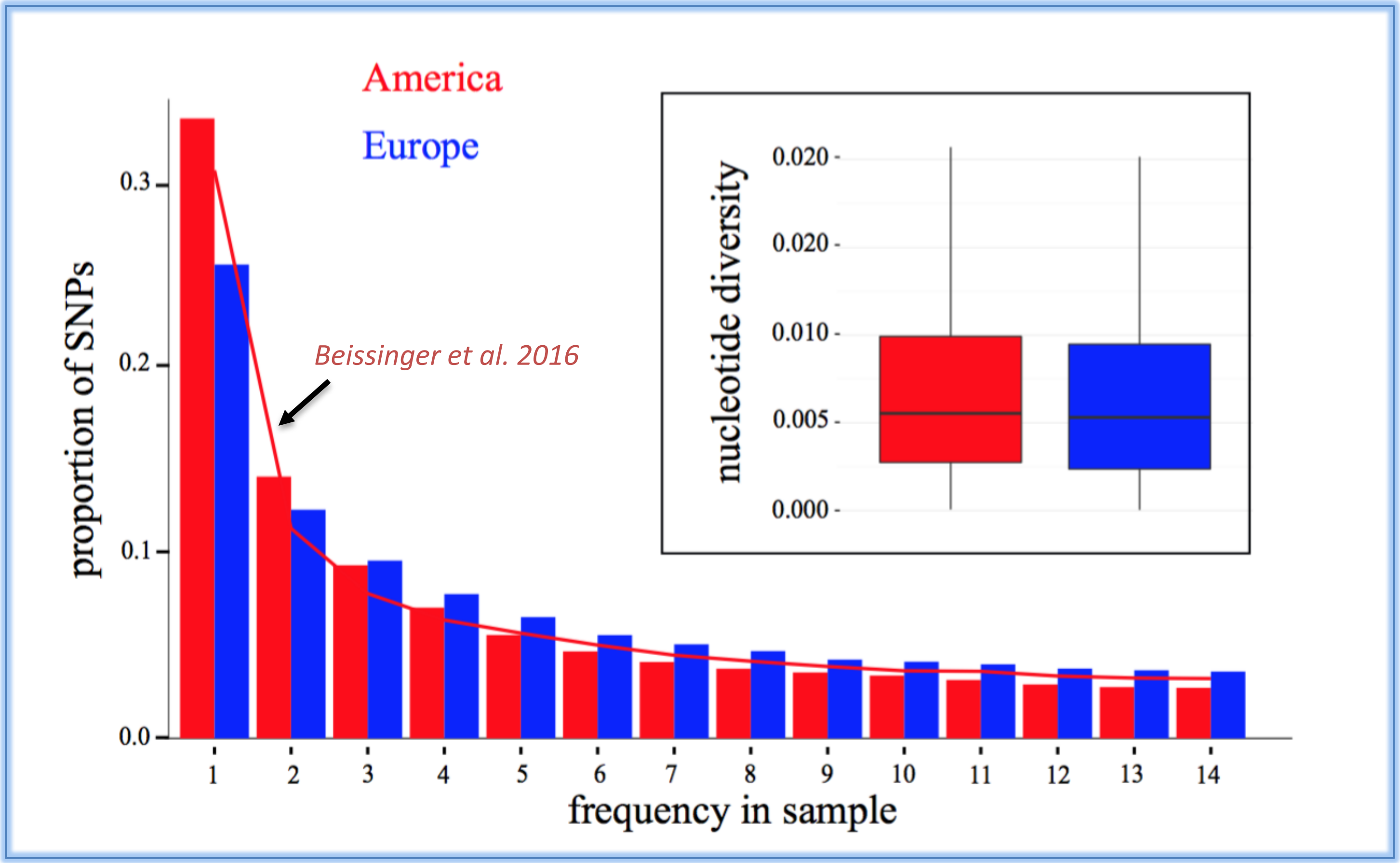
2 independent introductions, 3 admixed groups

# Proposed scenario of European introduction



Admixture is an important contributor of adaptation to mid-latitudes

# Footprints of European introduction



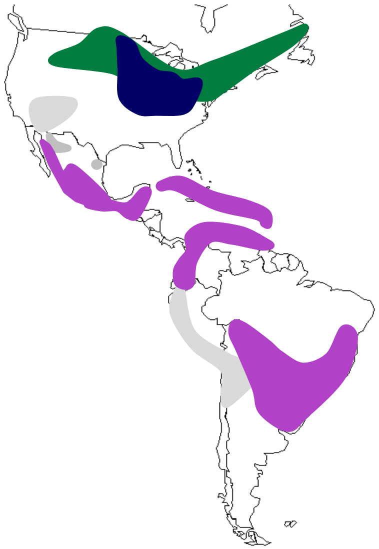
Modest footprints but:

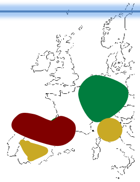
Significant loss of diversity (7%)

Loss of rare variants

# Targets of selection along latitudinal and longitudinal contrasts

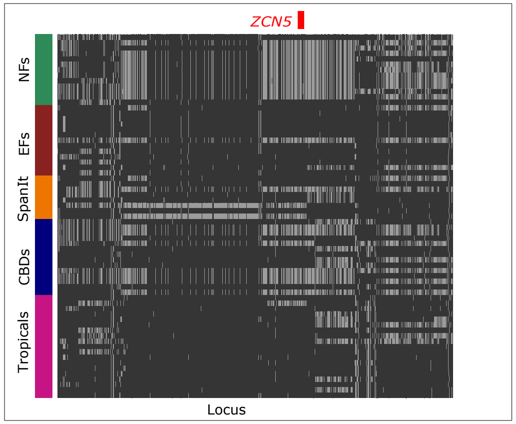
Combination of 3 statistics in genes, using non-genic windows as controls





| Latitudinal comparisons | | |
|---|---|---|
| NF | CBD | 163 (112/56) |
| CBD | Tropical | 46 (21/25) |
| NF | Tropical | 175 (124/61) |
| NF | EF | 201 (132/69) |
| EF | Span&It | 129 (50/81) |
| NF | Span&It | 219 (125/108) |
| **Longitudinal comparisons** | | |
| NFA | NFE | 147 (85/65) |
| CBD | EF | 122 (81/51) |
| Tropical | Span&It | 119 (13/106) |

*flowering candidate*



968 Differentially Selected candidates: more on average in latitudinal than in longitudinal contrasts, more in NFs than in the other groups

# Polygenic adaptation along gene networks

Adapted from *Daub et al. MBE 2013* :
294 networks
Sum of normalized P-values among genes of a given network
(P-values are computed from observed Fst and Fst of non-genic windows, network size is accounted for)

Drought tolerance
tetrapyrrole synthesis
ABA synthesis

Stress responses
cis- and trans-zeatin pathways

Cold tolerance
Putrescine pathway

Defense against pathogens
β-caryophylens

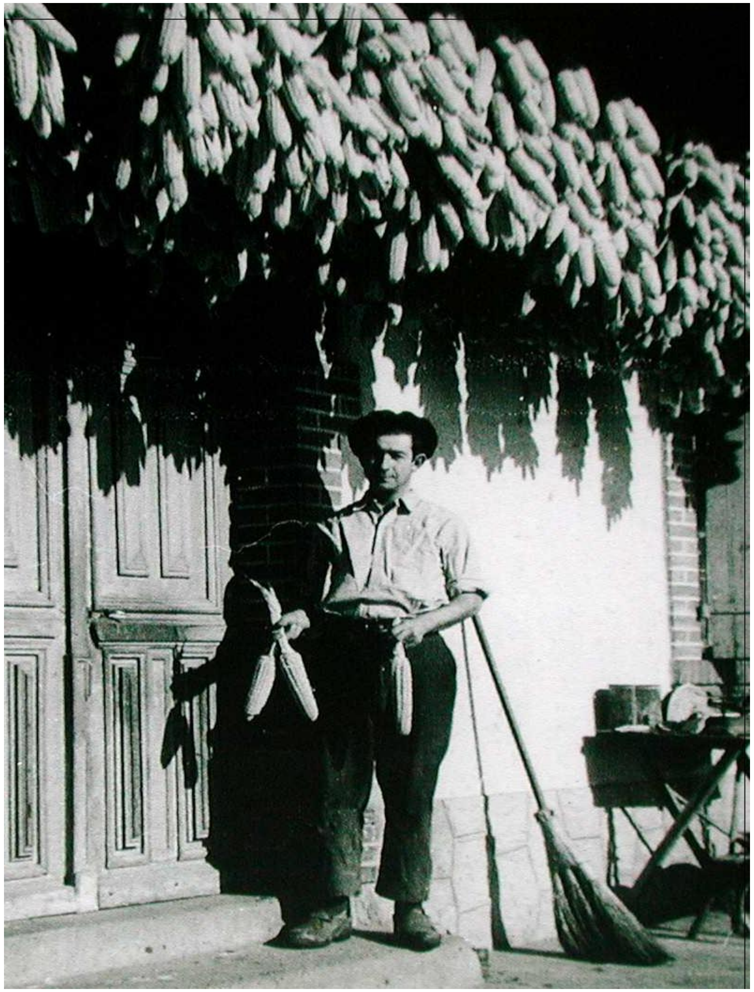
*Brandenburg et al. PLoS Genet 2017*

# Early European corn

One of the earliest european drawing of maize can be found in the translation of the book Materia medica de Dioscorides edition 1543 (Frankfurt)

Illustration taken from
*De historia stirpium* de Fuchs (1542)
with notes on:
the American origin of corn
the 4 grain colours painted on a single plant

# During the 18th century corn become a staple crop



# Outline

Drivers of maize adaptation:

- Short life cycle? Relation to genome size evolution

# What do we know at the intraspecific level in plants?

We know very little:

- <5% in *Hordeum maritimum* *Jakob et al. MBE 2004*

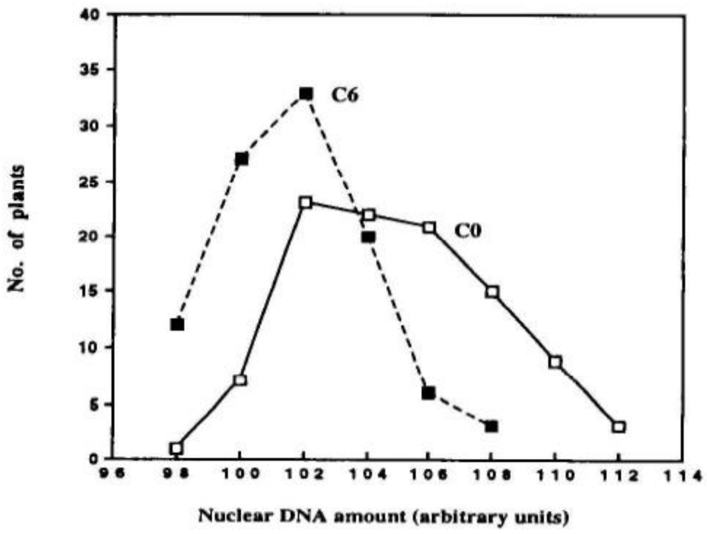
- 16.6% among *Festuca pallens* populations *Smarda et al. Annals Bot 2008*

- about 10% among *Arabidopsis* ecotypes *Long et al. Nat Gen 2013*
- about 30% of variation in maize and its wild relatives

Maize intraspecific variation makes it an interesting model for studying factors driving GS variation

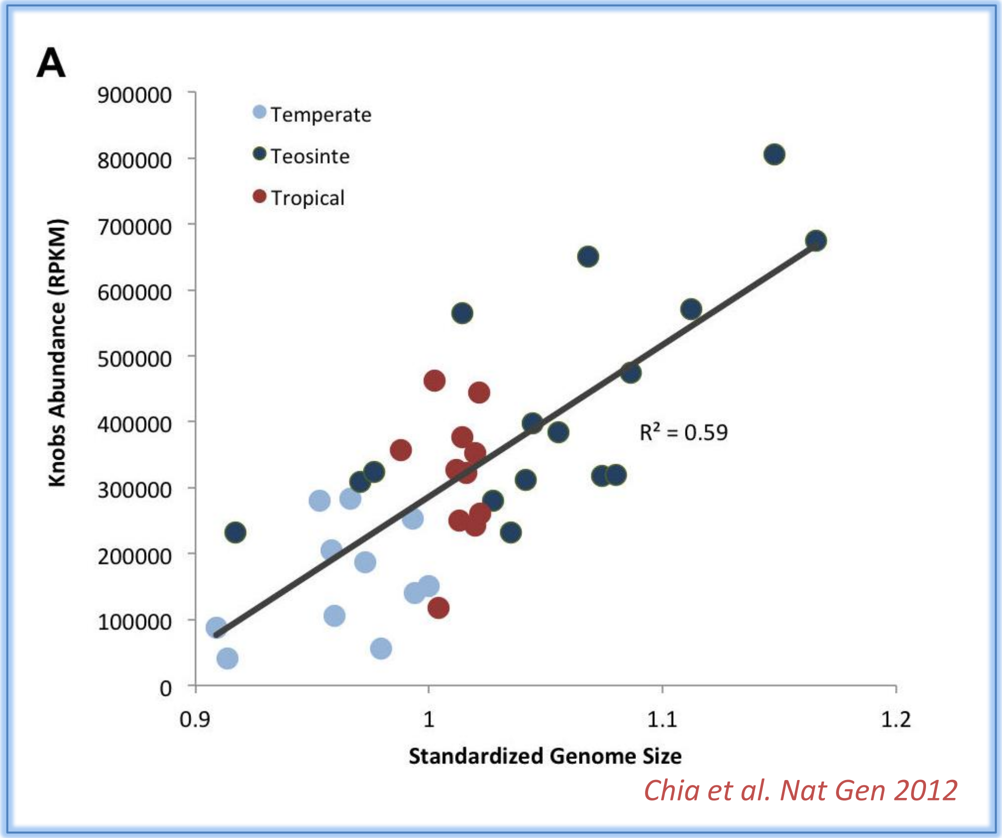
# Maize genome fluidity

Distribution of DNA content at G0
and after 6G of selection for earliness



*Rayburn et al. Plant Breeding 1994*

# Genome size correlates with knob content



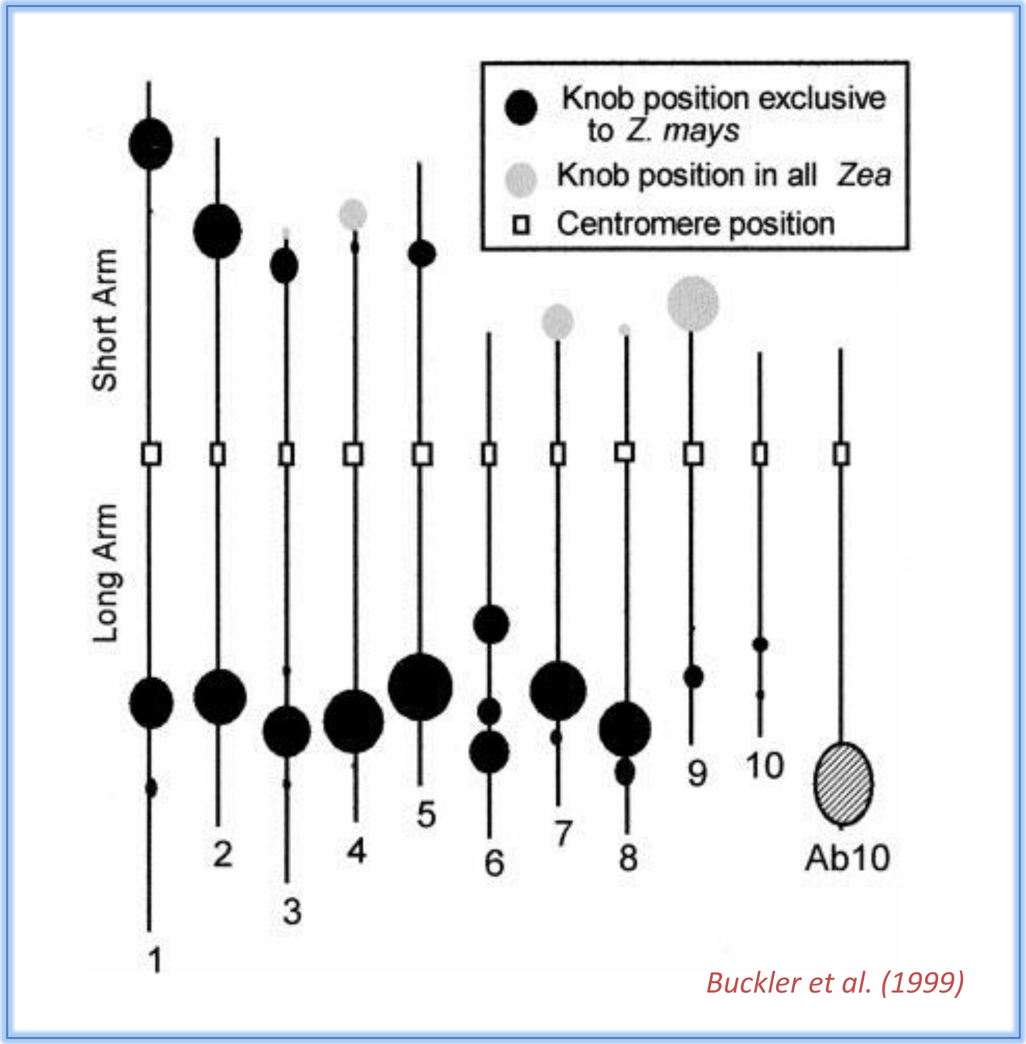
*Chia et al. Nat Gen 2012*

Genome size correlates with knob content (P-value=0.00046)

GS teosintes > GS maize inbred lines
GS tropicals > GS temperate

➔ GS may be adaptive (seems to correlate with environmental variables)

# Knobs in *Zea*



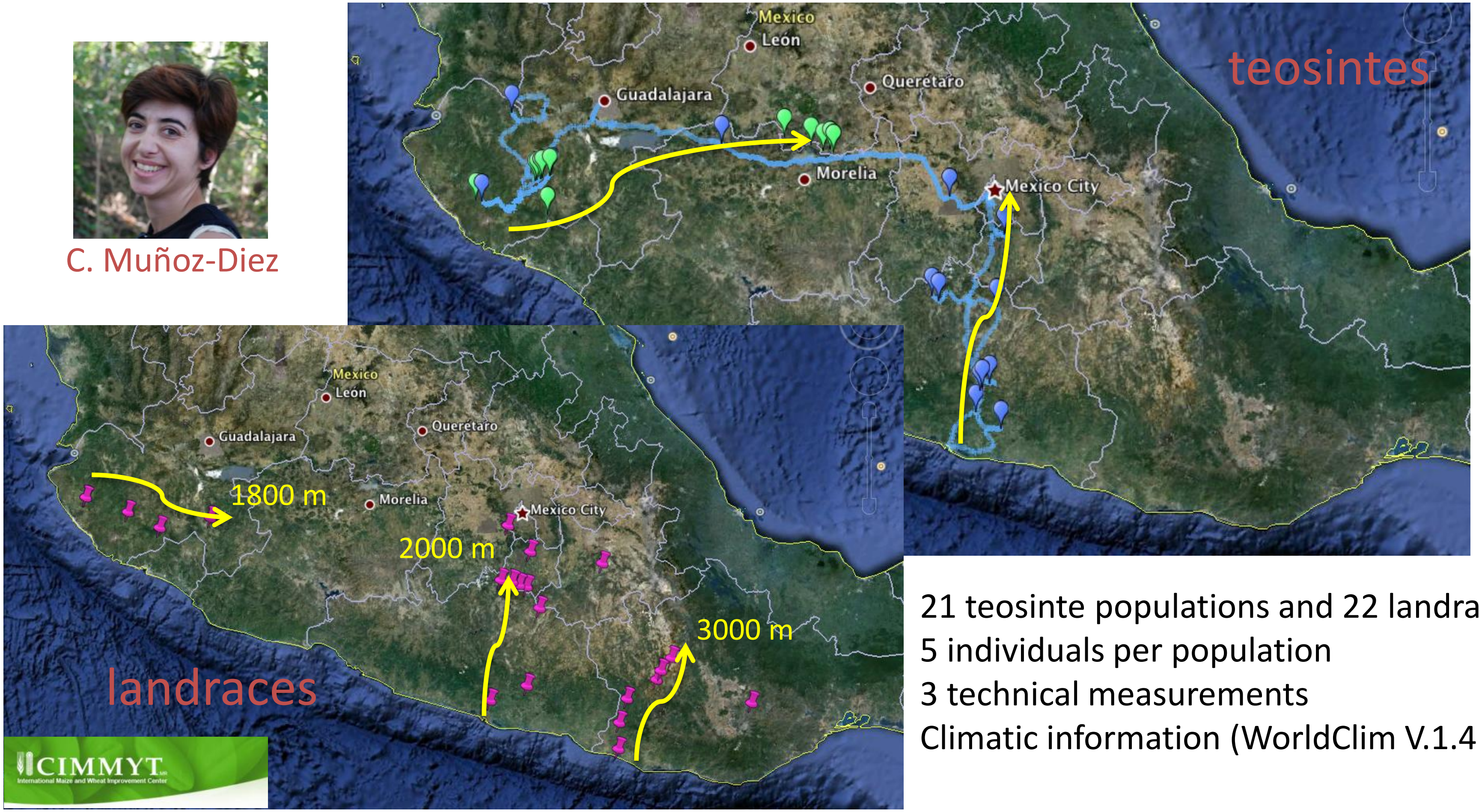
*Buckler et al. (1999)*

Area of circle is proportional
to size and frequency of the knobs

Knobs

- vary in size and frequency
- are constituted by 180-bp 360-bp tandem repeats and transposable elements

# Correlation between genome size and ecological variables

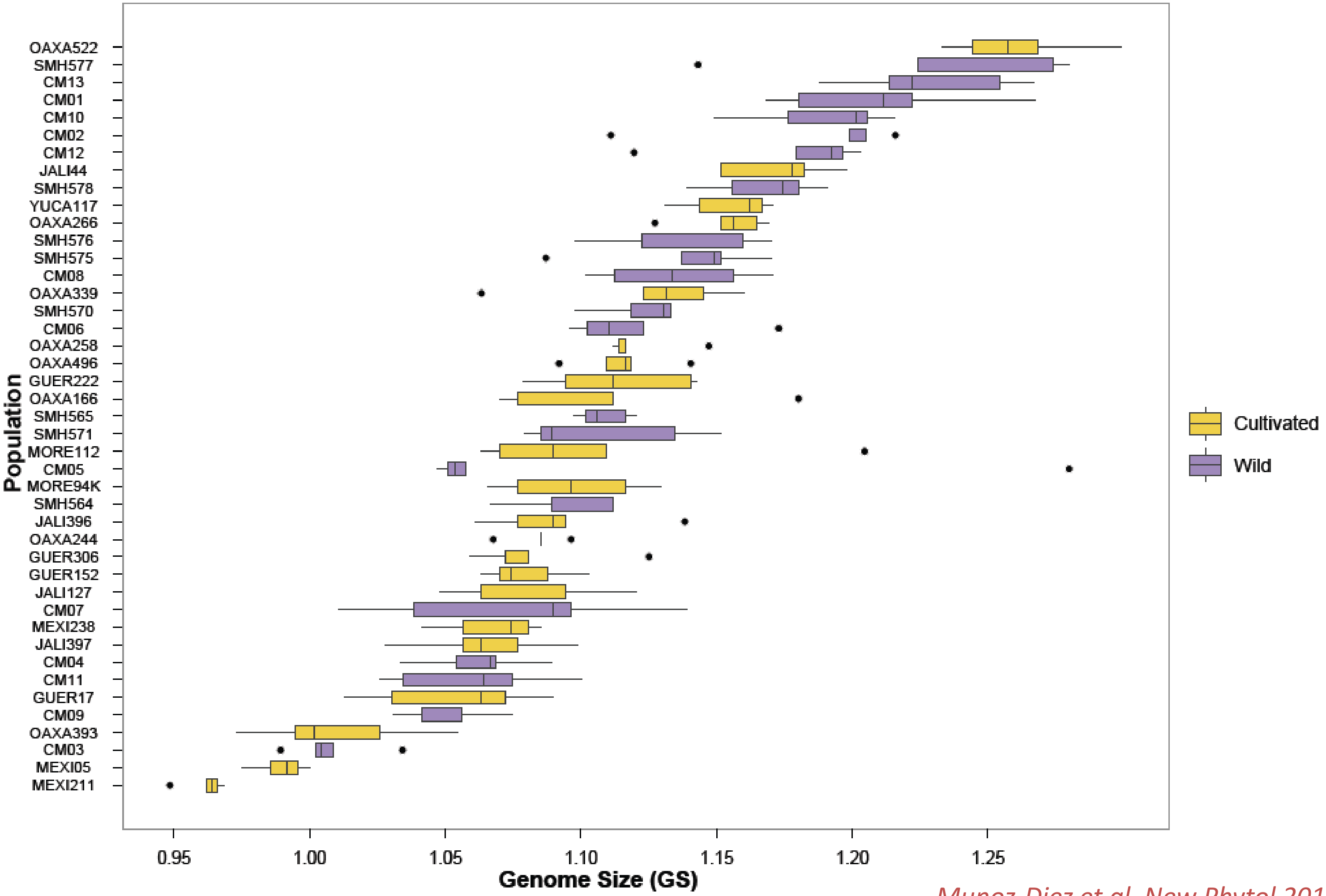


C. Muñoz-Diez

21 teosinte populations and 22 landraces
5 individuals per population
3 technical measurements
Climatic information (WorldClim V.1.4)

Conception Muñoz-Diez, Esteban Meca, Brandon Gaut, UC Irvine, USA
Salvador Montes-Hernandez, INIFAP Celaya, Mexico
Enrique Scheinvar, Luis Eguiarte, UNAM, Mexico

# Estimating genome size in teosintes and landraces

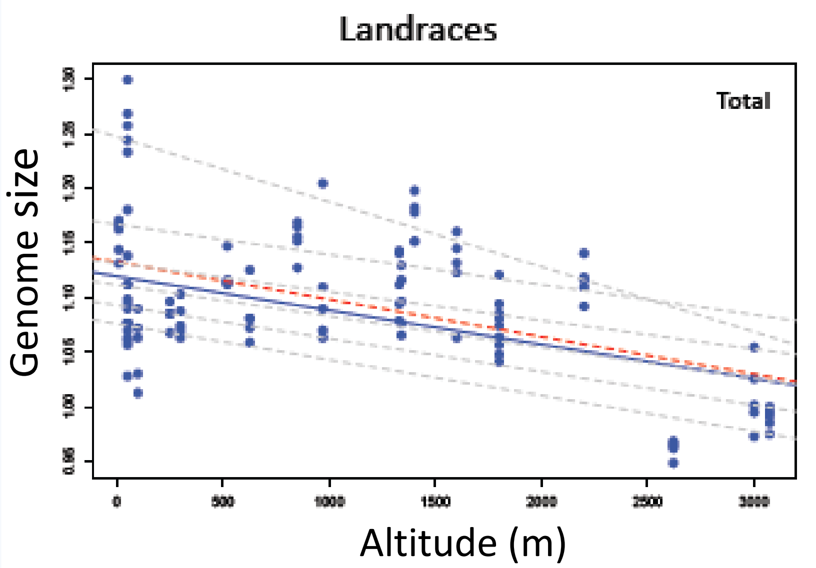


*Munoz-Diez et al. New Phytol 2013*

Genome size is significantly larger in teosintes than in maize (by few percent)

Proportion of variance explained by variation among populations (≈71%) > within (≈20%) > subspecies (≈2%)

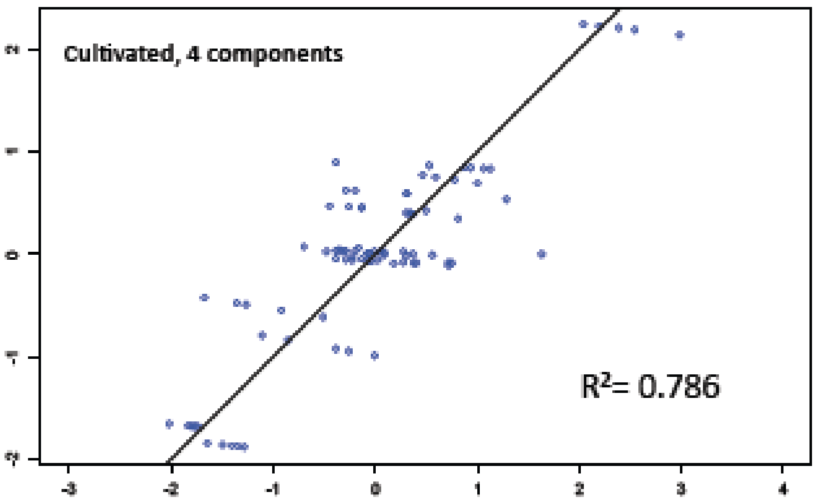
# Is genome size variation adaptive in *Zea mays* ?



*Quantile regression with median fit and least square estimate of the mean*

In maize, GS correlates negatively with altitude

And geographical components (altitude, longitude, latitude) are good Predictors of GS



*observed and predicted GS from partial least square regression*

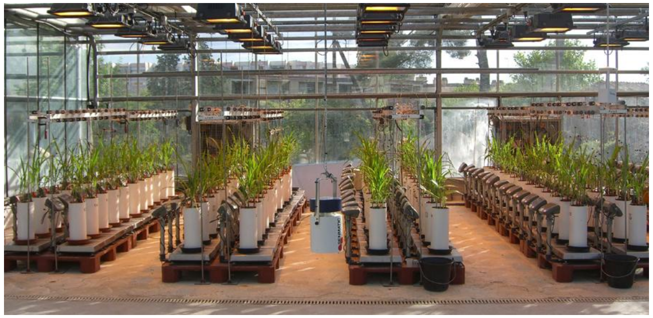
Phenotypes that affect GS are likely complex, we investigated one of them: Leaf Elongation Rate

# Measuring growth with LERmax

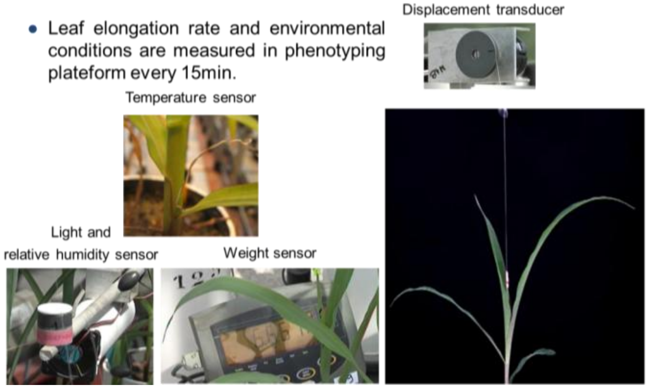
J. Melique

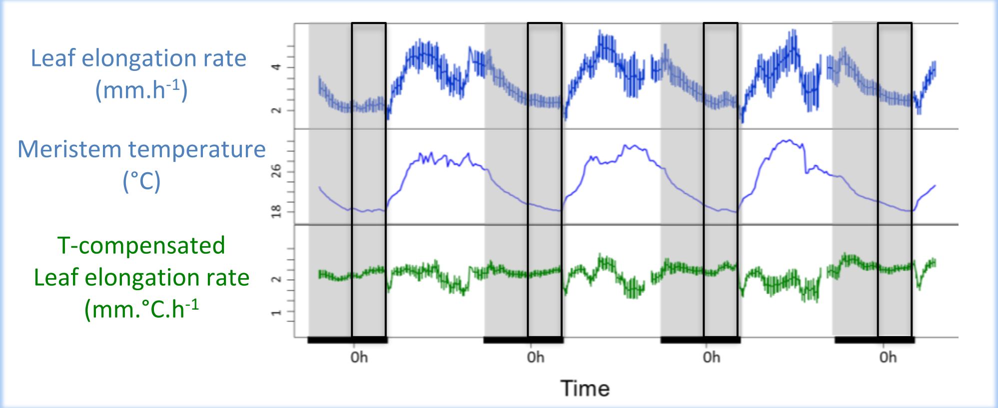
Claude Welcker, INRA Montpellier, France
Francois Tardieu, INRA Montpellier, France
Domenica Manicacci, Le Moulon, France

Measuring Leaf Elongation Rate at Phenodyn (http:montpellier.inra.fr/bioweb/phenodyn)



- Leaf elongation rate and environmental conditions are measured in phenotyping plateform every 15min.

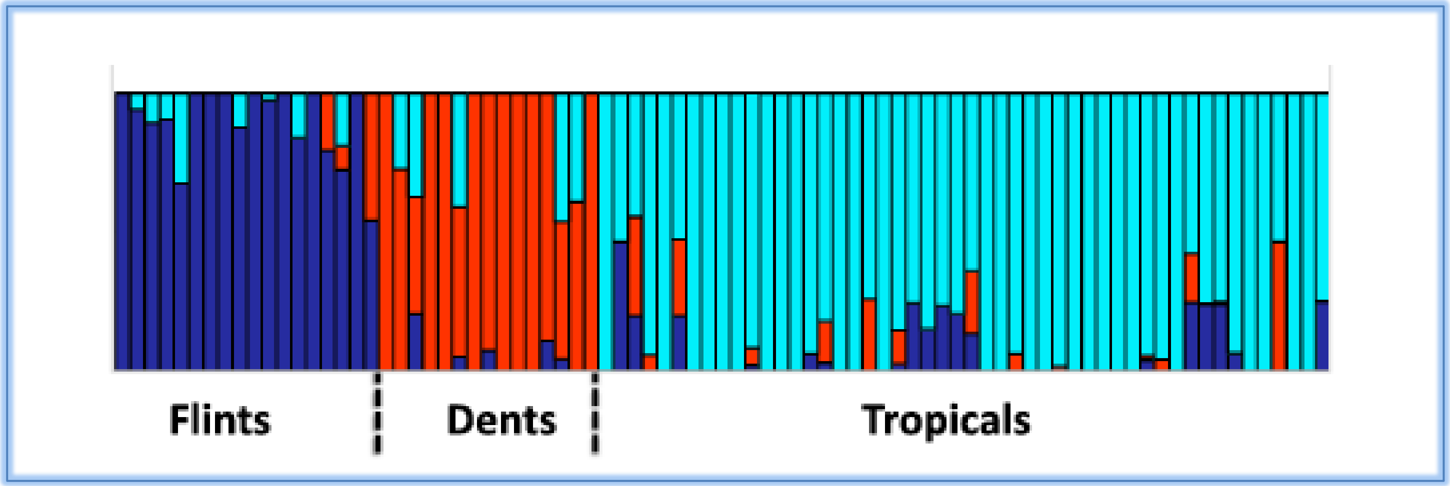




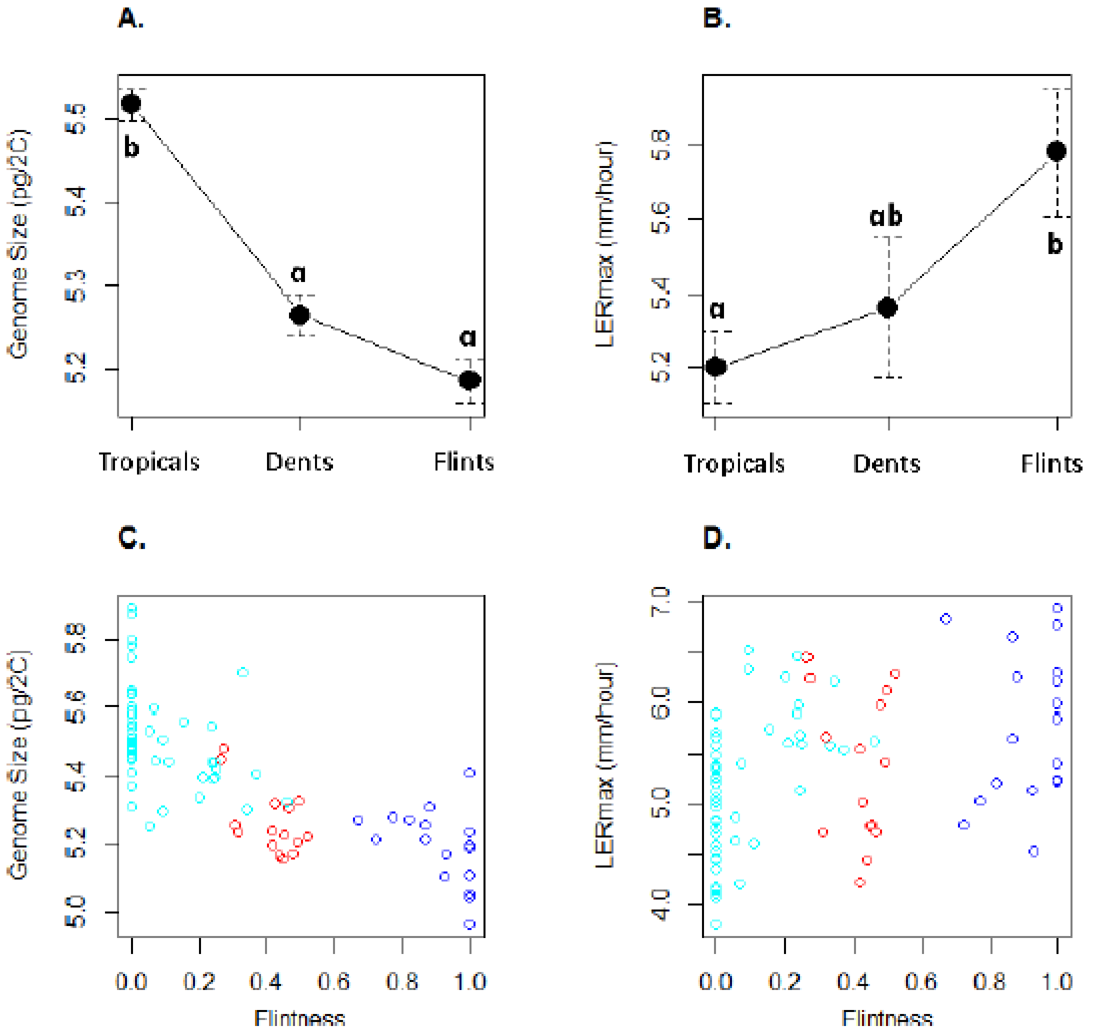
- Leaf growth fluctuates with °C
- In thermal time, a plateau is reached between 0-4h= LERmax
- LERmax measures the intrinsic ability of leaf to grow
- It is highly heritable

*Sadok et al. Plant Cell Env 2007; Dignat et al. Plant Cell Env 2013*

# Sample: 83 maize inbred lines structured in 3 genetic groups



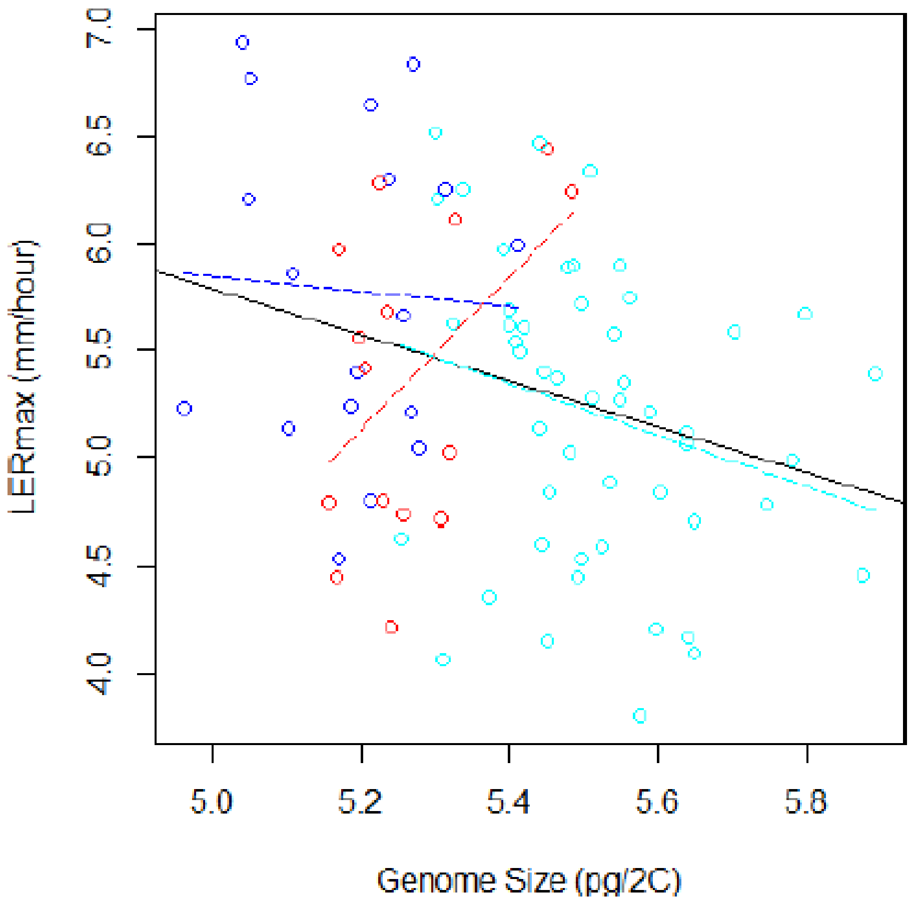
# Genome size and LER max by genetic group



*R=-0.77, P.value=2.1 $10^{-17}$*

*R=0.40, P.value=2.1 $10^{-4}$*

# Relationship between genome size and LER max



*R=-0.29, P.value=0.008*
*Group effect significant, P. value=0.012*
*After correcting for Structure, P. value=NS*

Correlation mainly driven by among-group variation

*Tenaillon et al. PeerJ 2016*

# Repeated patterns in Meso- and South America

*Bilinski et al. bioRxiv, 2017 – Ross-Ibarra lab, UC Davis*

Negative correlations between GS and altitude, knob repeat and TE content in Meso- and South-America (-108 and -154 Kb/meter) in maize landraces

Negative relationship between GS and cell production rate (inferred from LER) within a *mexicana* population,

Negative relationship between flowering time and cell production in the meristem in maize lines *Leiboff et al. Front Plant Science 2016*

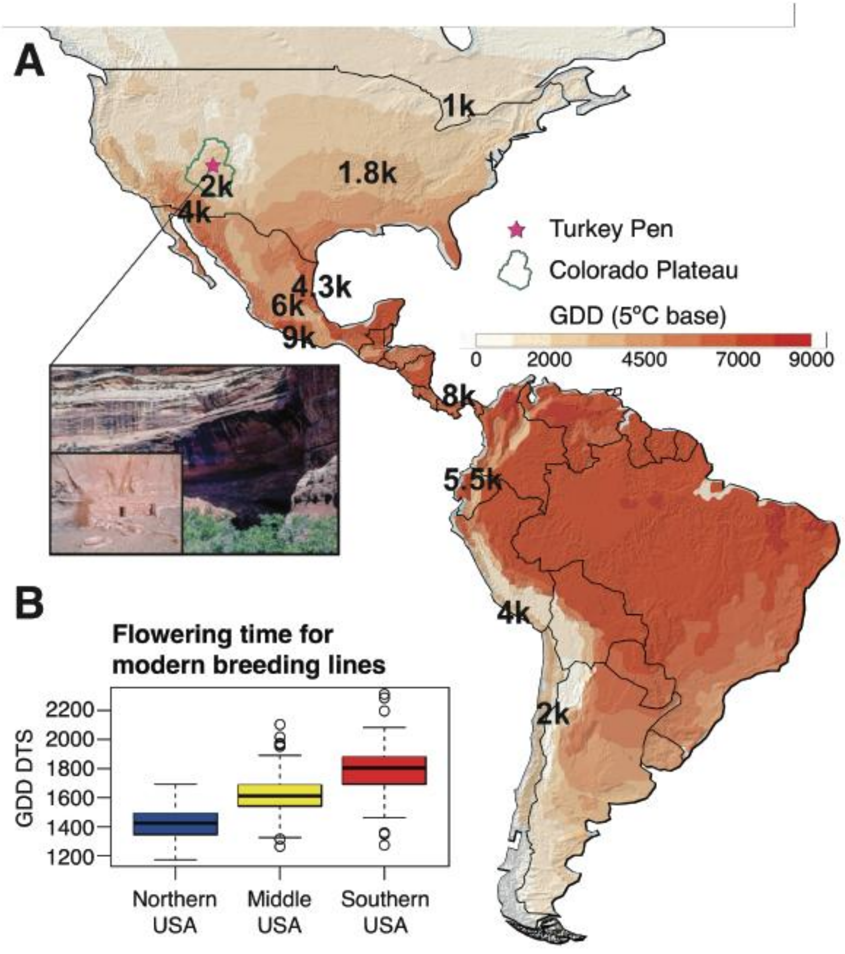
Drivers of maize adaptation:

- Flowering time

# Flowering time adjustment: a major component of maize diffusion

Maize is originally a tropical plant
sensitive to photoperiod
flowers under short days (< 13hours of daylight).
Variation in daylength indicates a transition
between the dry and wet season in the Tropics.

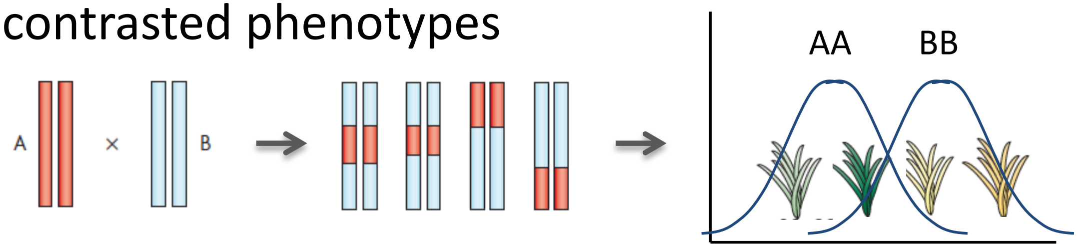
In temperate regions, the days are longer
in the summer and temperatures lower.
Selection of maize has triggered
a loss of sensitivity to photoperiod,
early flowering to lengthen the
growing season.



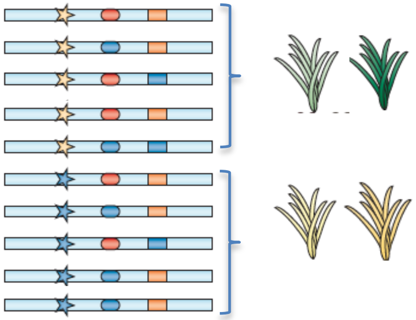
*Swarts et al. Science 2017*

# Linking phenotypic to genotypic variation

- QTL mapping: Assessing genetic differences in progenies of crosses between two (few) parents with contrasted phenotypes

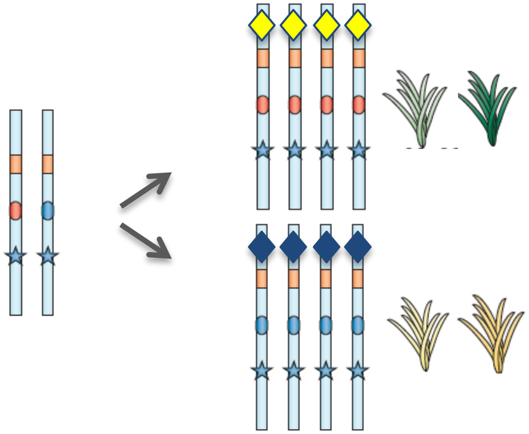


- Genome Wide Association mapping (GWA): Statistically associating genotypic to phenotypic variation



- Selection experiments: Moving phenotypes artificially in two directions from highly inbred material

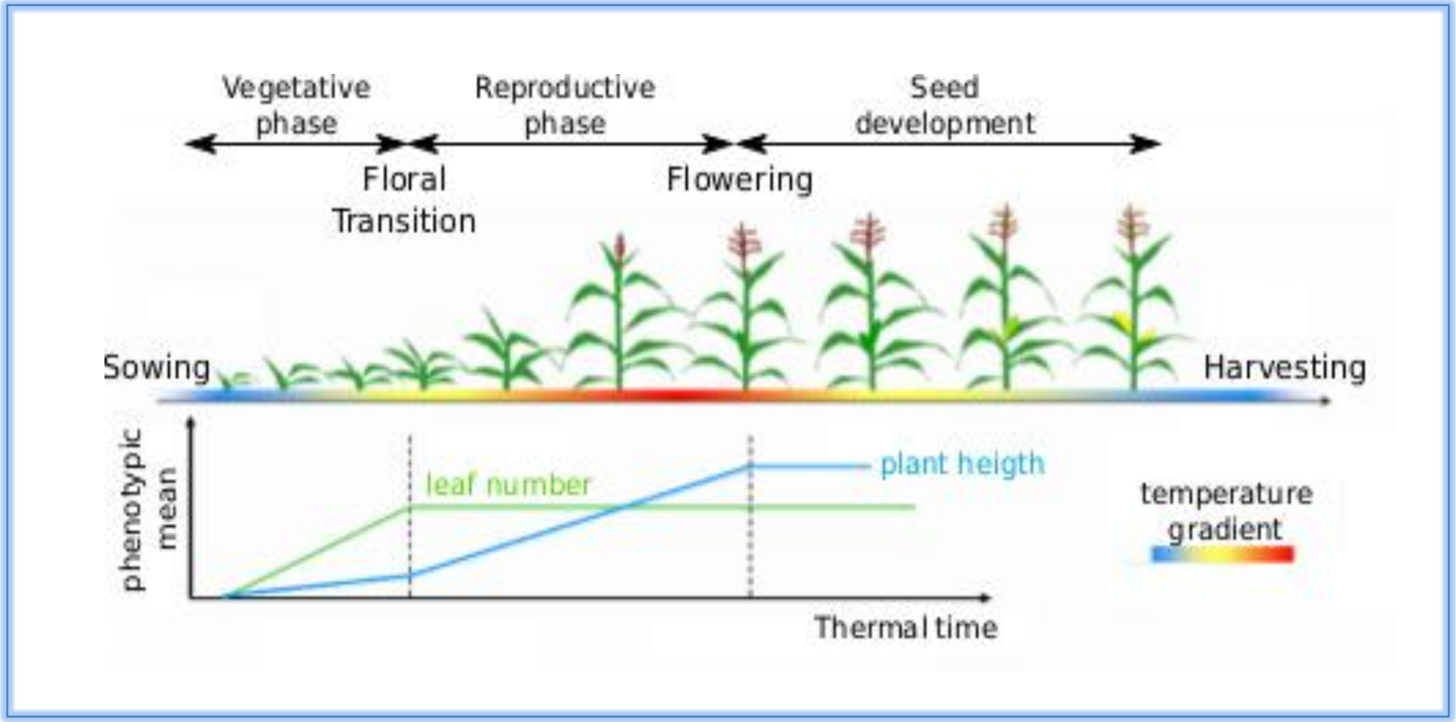
By better controlling the genetic background, the history of the population and the number of traits under selection



*Takeda and Matsuoka Nat Rev Genet 2008*
*Bergelson and Roux Nat Rev Genet 2010*

# Maize flowering time

Floral transition marks the transition between vegetative and reproductive phase, and contributes to determine flowering time

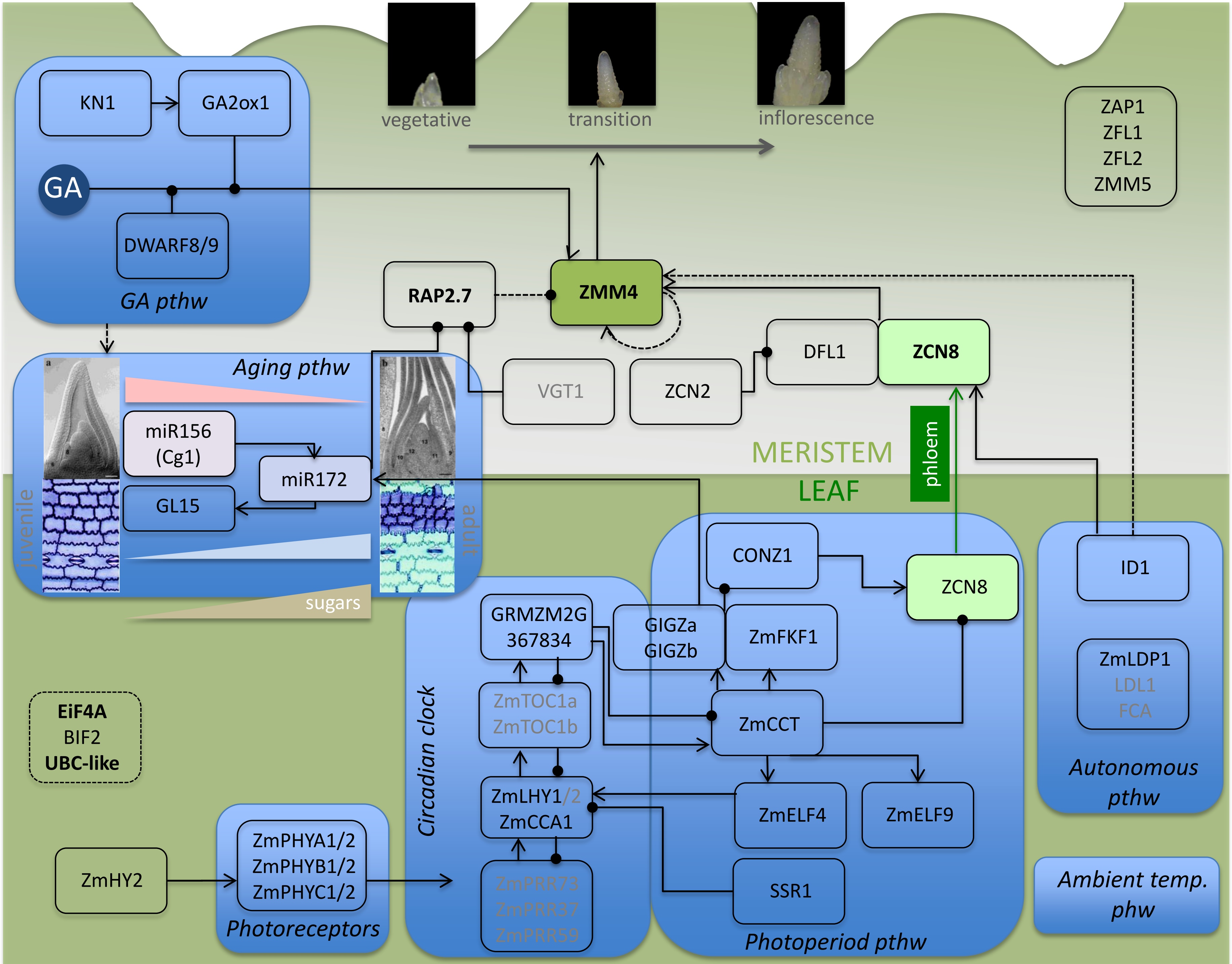


E. Durand

Christine Dillmann, GQE-Le Moulon, France

Adrienne Ressayre, GQE-Le Moulon, France
Elodie Marchadier, IPS2, France
Martine Le Guilloux, GQE-Le Moulon, France
Aurélie Bourgais, GQE-Le Moulon, France
Hélène Corti, GQE-Le Moulon, France

LabEx BASC
Biodiversity, Agroecosystems, Society, Climate



vegetative
transition
inflorescence
ZAP1
ZFL1
ZFL2
ZMM5
KN1
GA2ox1
GA
DWARF8/9
GA pthw
RAP2.7
ZMM4
DFL1
ZCN8
Aging pthw
VGT1
ZCN2
miR156
(Cg1)
miR172
GL15
juvenile
adult
MERISTEM
LEAF
phloem
sugars
CONZ1
ZCN8
ID1
GRMZM2G
367834
GIGZa
GIGZb
ZmFKF1
ZmLDP1
LDL1
FCA
EiF4A
BIF2
UBC-like
Circadian clock
ZmTOC1a
ZmTOC1b
ZmCCT
Autonomous
pthw
ZmLHY1/2
ZmCCA1
ZmELF4
ZmELF9
ZmHY2
ZmPHYA1/2
ZmPHYB1/2
ZmPHYC1/2
ZmPRR73
ZmPRR37
ZmPRR59
SSR1
Ambient temp.
phw
Photoreceptors
Photoperiod pthw

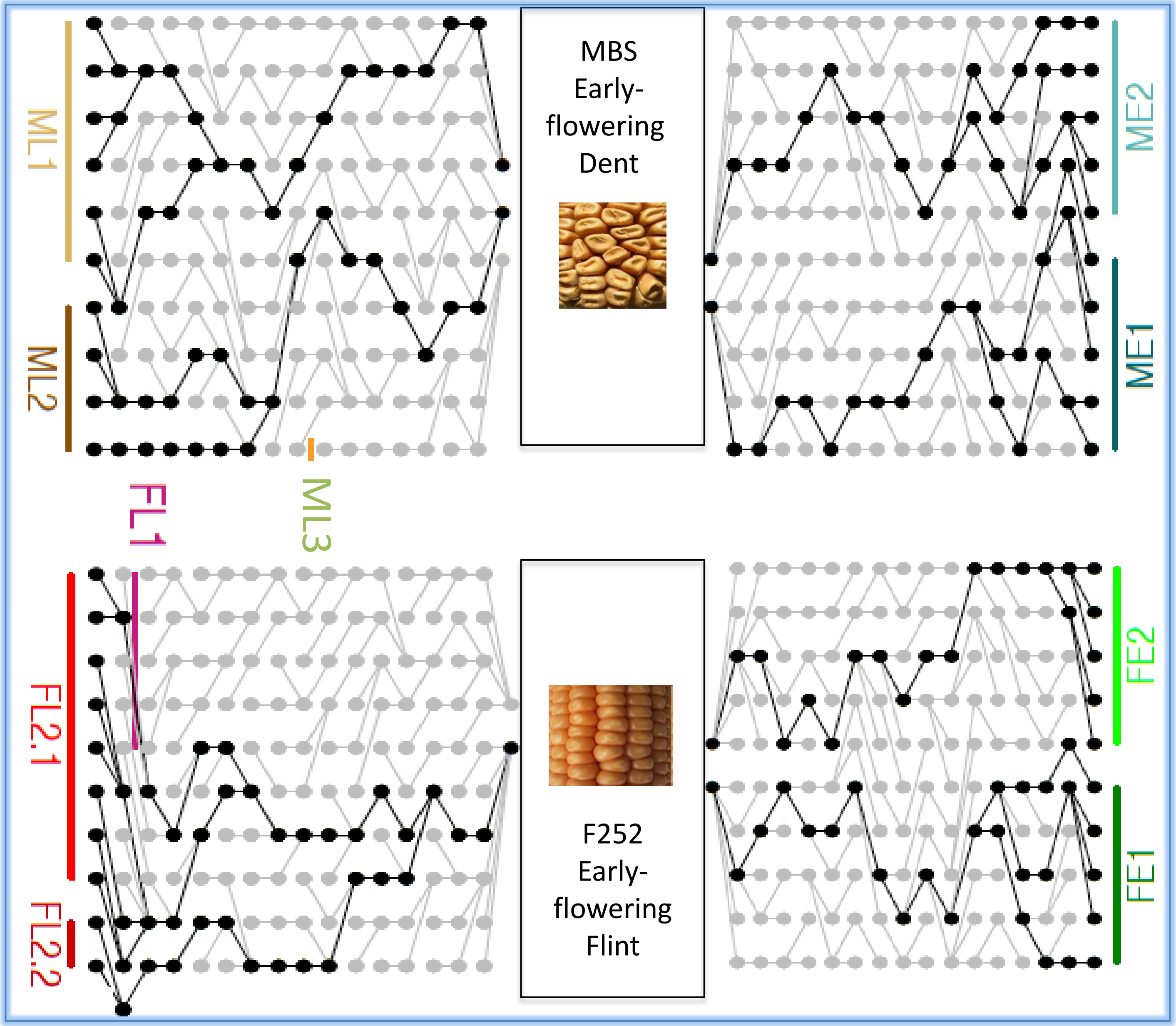
# 16 years of divergent selection for flowering time

**Every generation:**

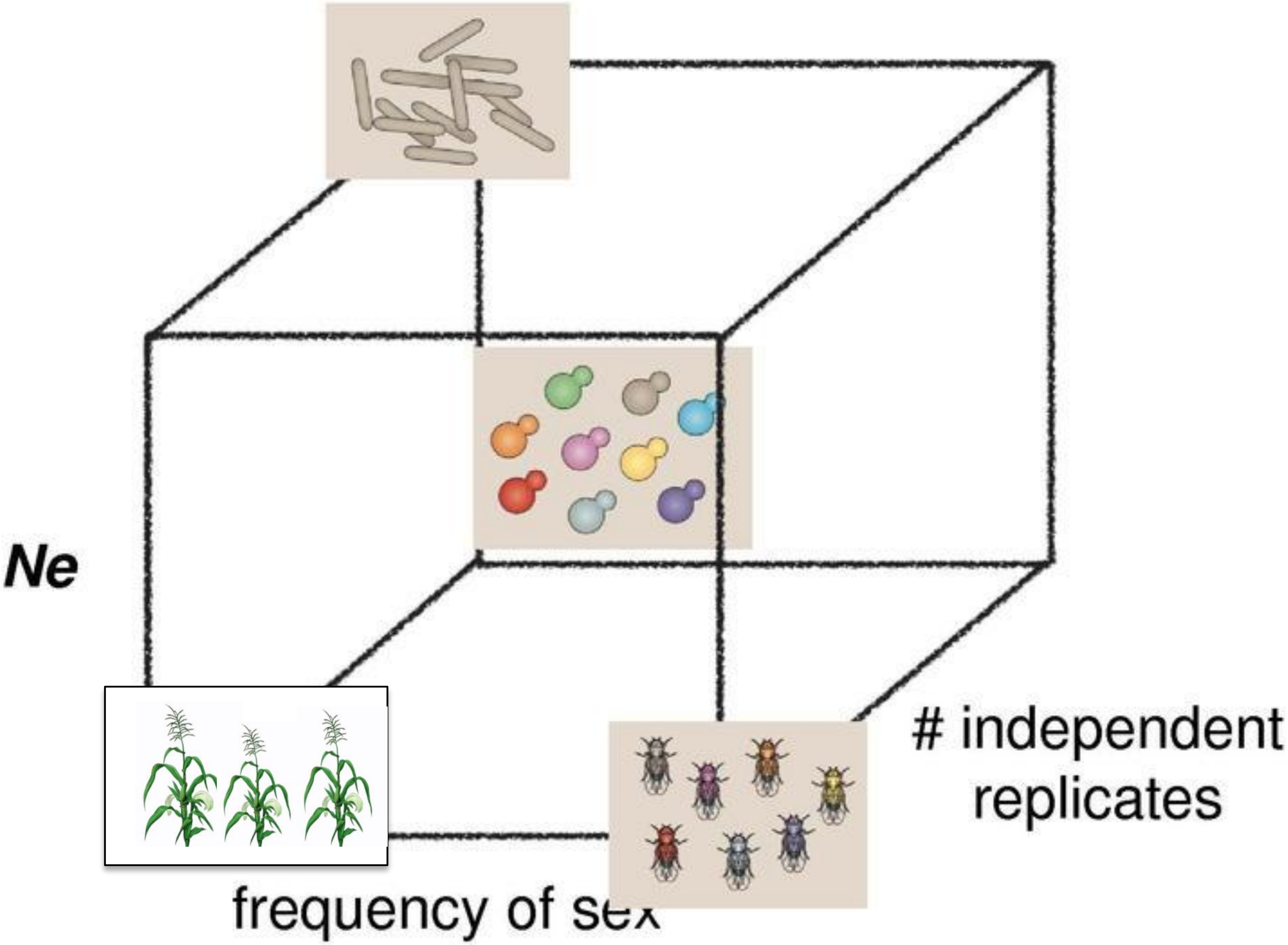
Selection of **10 Earliest** and **10 Latest** genotypes

**Selfing**

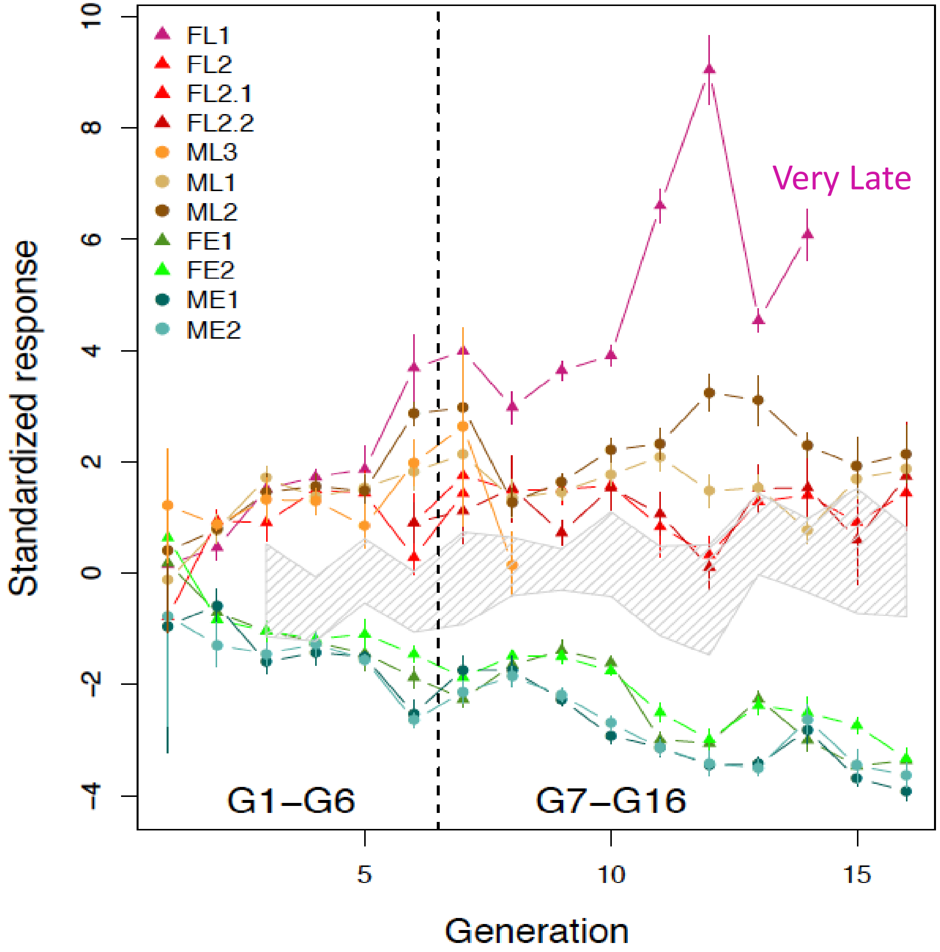
**Phenotypic evaluation** of 100 plants/genotype = 2000 plants total per selection experiment



# Going back to Molly Burke slide



# Characterization of the response to selection from G1 to G16



Response expressed in degree days with sd around the mean, max/min for the controls

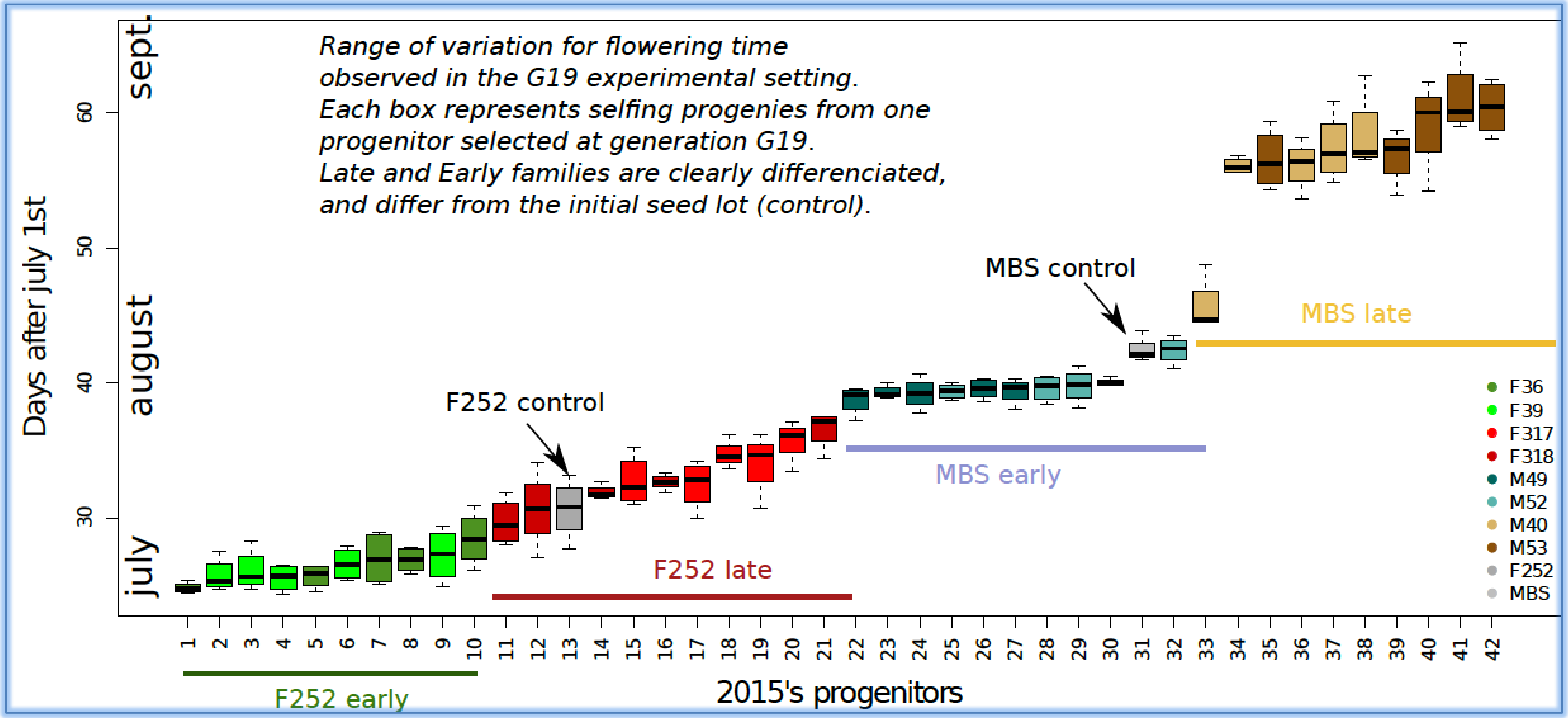
*12 days time lag between Early and Late genotypes*

Sustainable, strong and significant response to selection from G 1-6 (stronger in Late)

Significant in 6 over 8 families from generation 7-16, continuous in Early, much less so in Late

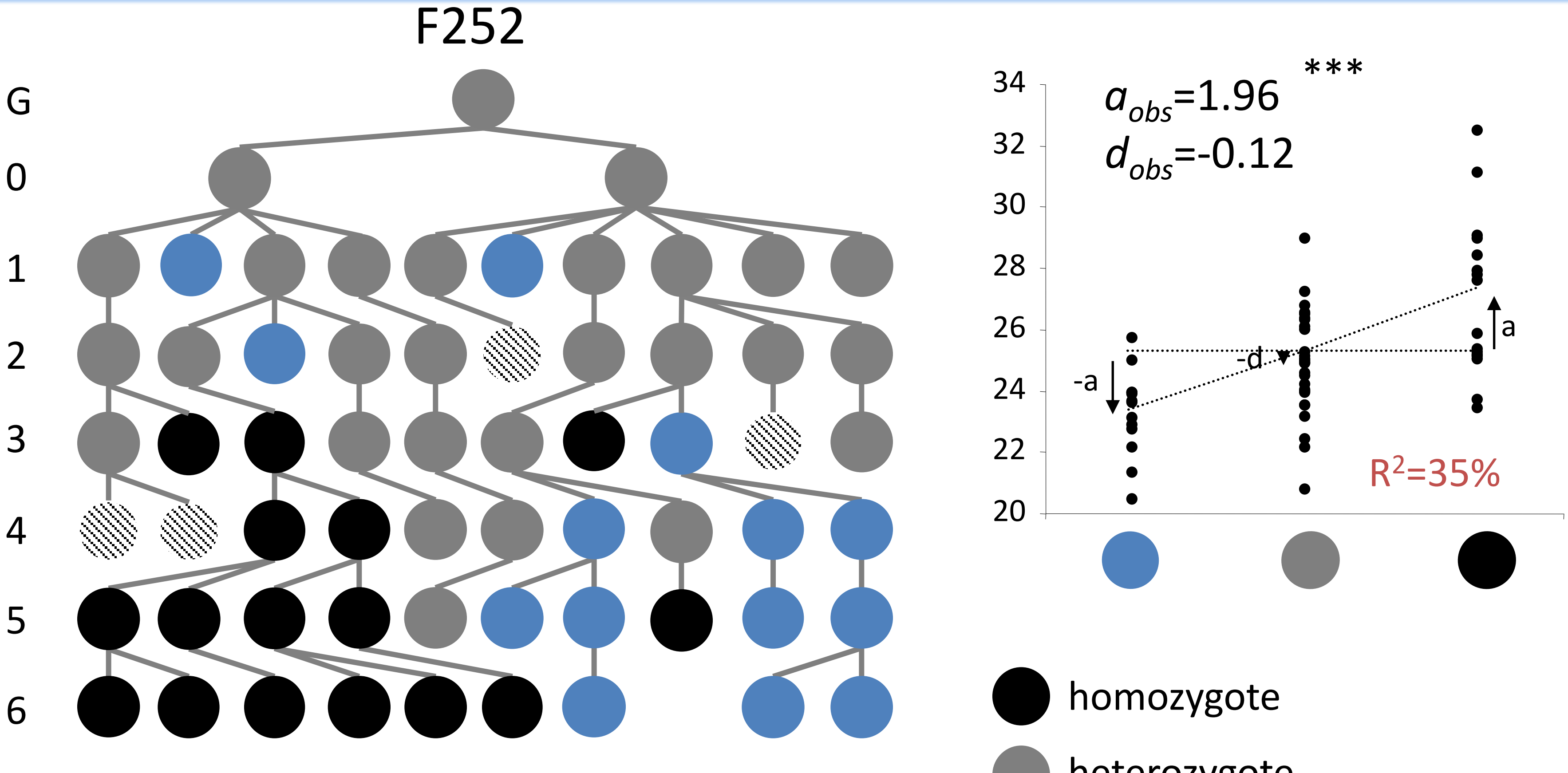
*Durand et al. BMC Evol. Biol 2010*

# At G19: up to 3 weeks difference between Early and Late



*Range of variation for flowering time observed in the G19 experimental setting. Each box represents selfing progenies from one progenitor selected at generation G19. Late and Early families are clearly differenciated, and differ from the initial seed lot (control).*

# A major locus associated with flowering time variation in Late F252



● homozygote

● heterozygote

● homozygote

*Durand et al. BMC Evol. Biol. 2010*

Polymorphism present as standing variation

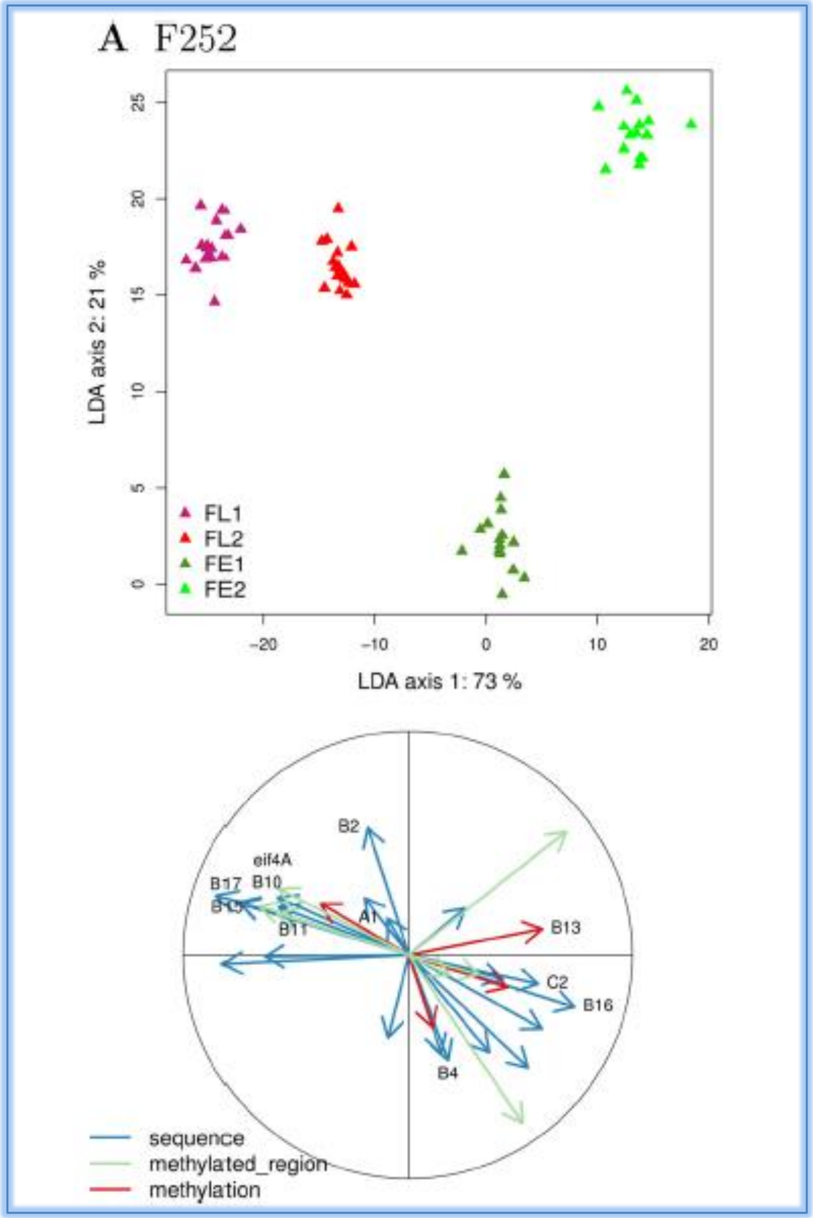
Locus may explain up to 35% of the variation for flowering time in the Late F252 population

# Other determinants: M-SAP and AFLP screening from G0-G6

24 markers total (highly inbred material) discriminating Early and Late subfamilies at G7

16 of these are sequence based-polymorphisms
8 are methylation-based polymorphisms

All markers were segregating in the original seed lot, no *de novo* mutations



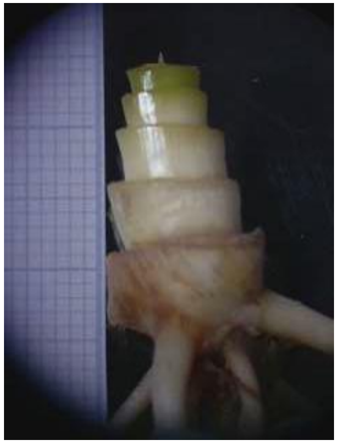
Durand et al. BMC Evol. Biol 2015

# In progress: Samples preparation (G13 genotypes)

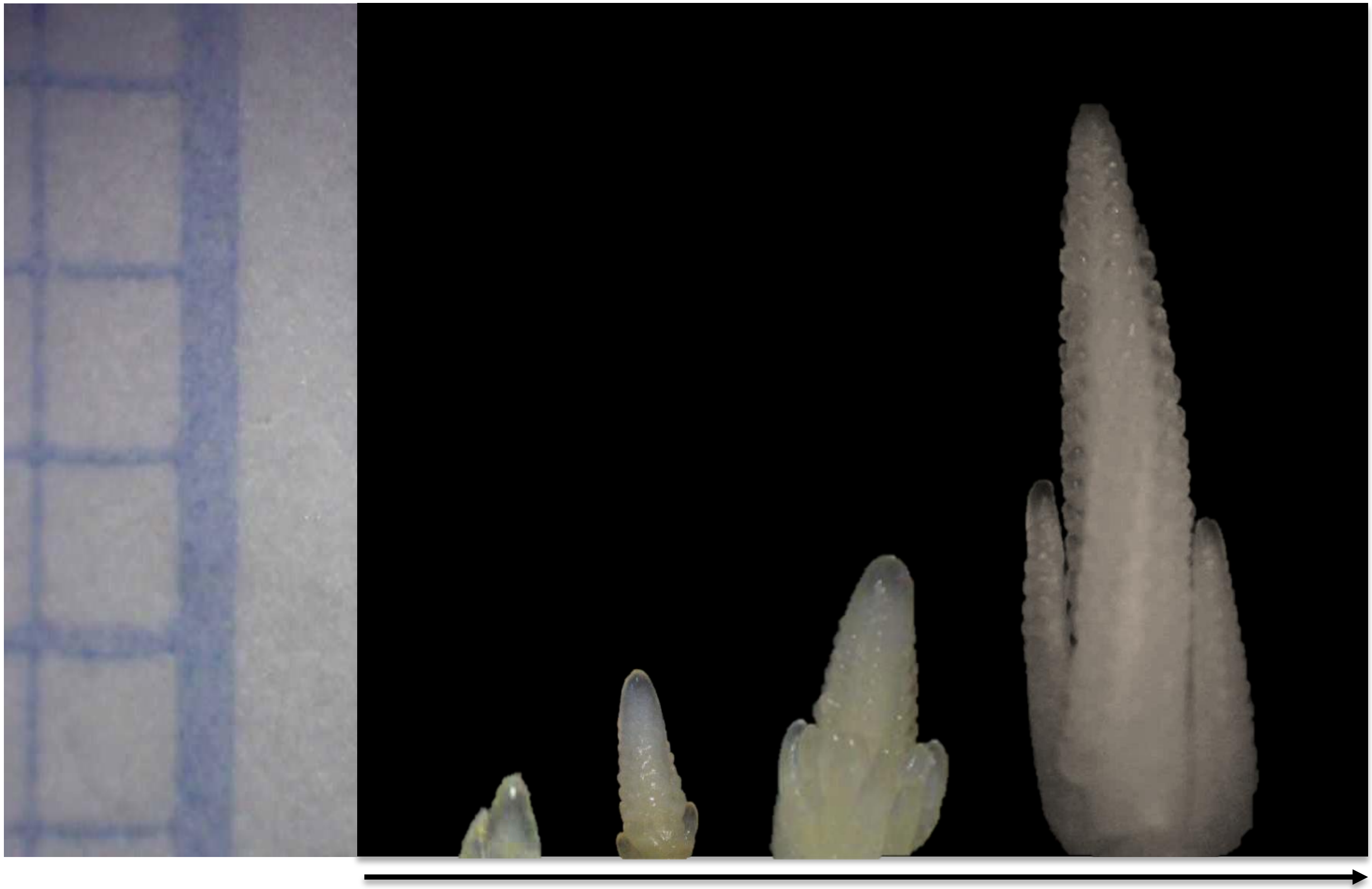

june/july in sunny Paris

Sampling, leaf number and meristem stage determination


Plant at L8



# Meristem status



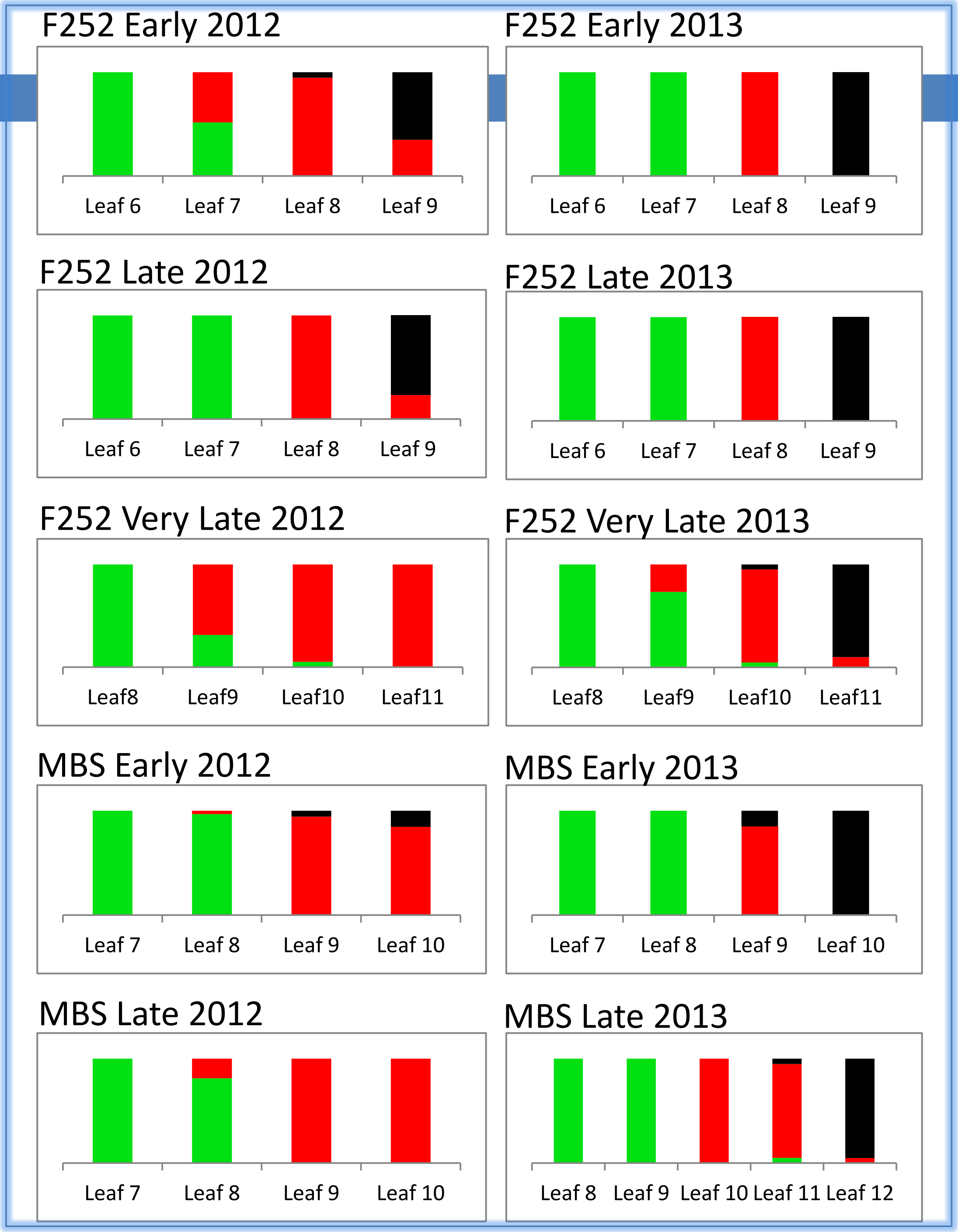
*time*

🟩 Vegetative

🟥 Transitioning

⬛ Reproductive

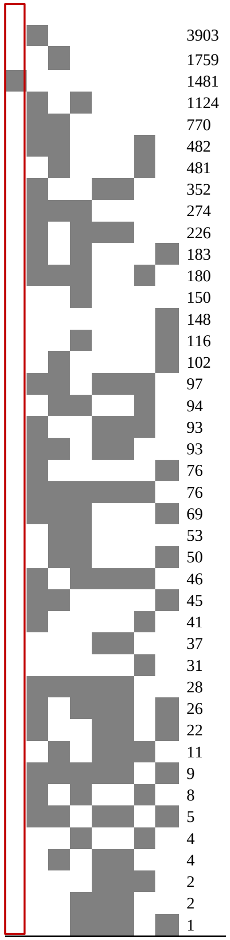
# Collected samples



Proportion of meristems stages in pools of 20 to 35 meristems

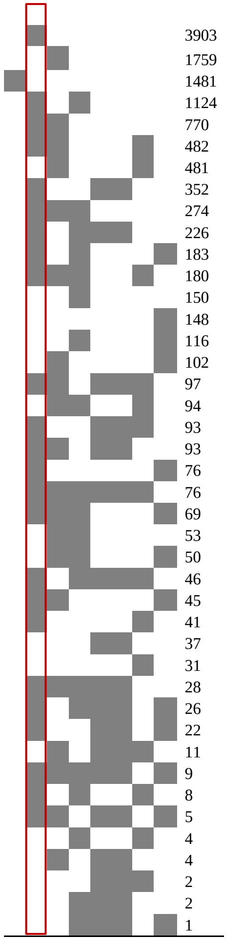
# Defining DE genes from contrasts

Year



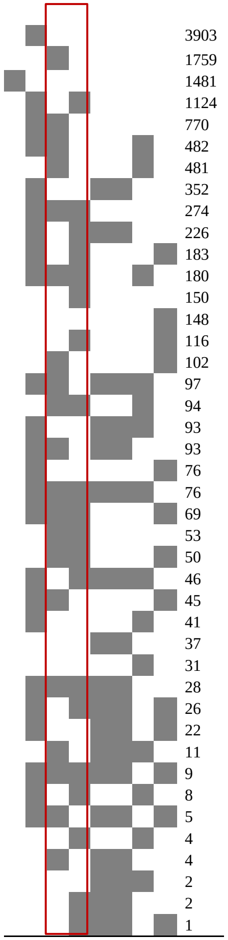
# Defining DE genes from contrasts

## Inbred Line



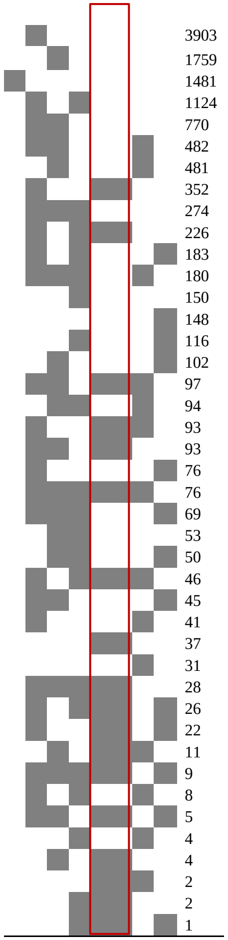
# Defining DE genes from contrasts

## Status in F252 and MBS



# Defining DE genes from contrasts

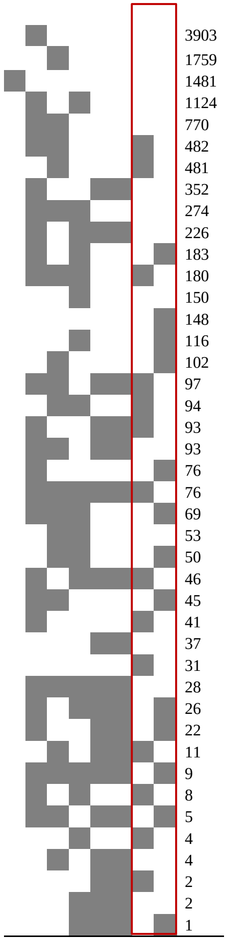
Early vs Late/VeryLate in F252 and MBS=**DE genes involved in the Early vs Late response**



# Defining DE genes from contrasts

Status x Genotype and Status x Genotype differing in Early vs Late/VeryLate in F252

**=DE genes involved in the Early vs Late response in F252**



# DE genes involved in the response to Early Late selection

39,066 genes tested
12,754 DE genes (33%)

1481 with Year effect only
2451 DE genes that differ between Early and Late genotypes (19%)

SELECTION

2120 in F252
446 in MBS
115 common (P.value=9.14 $10^{-6}$)

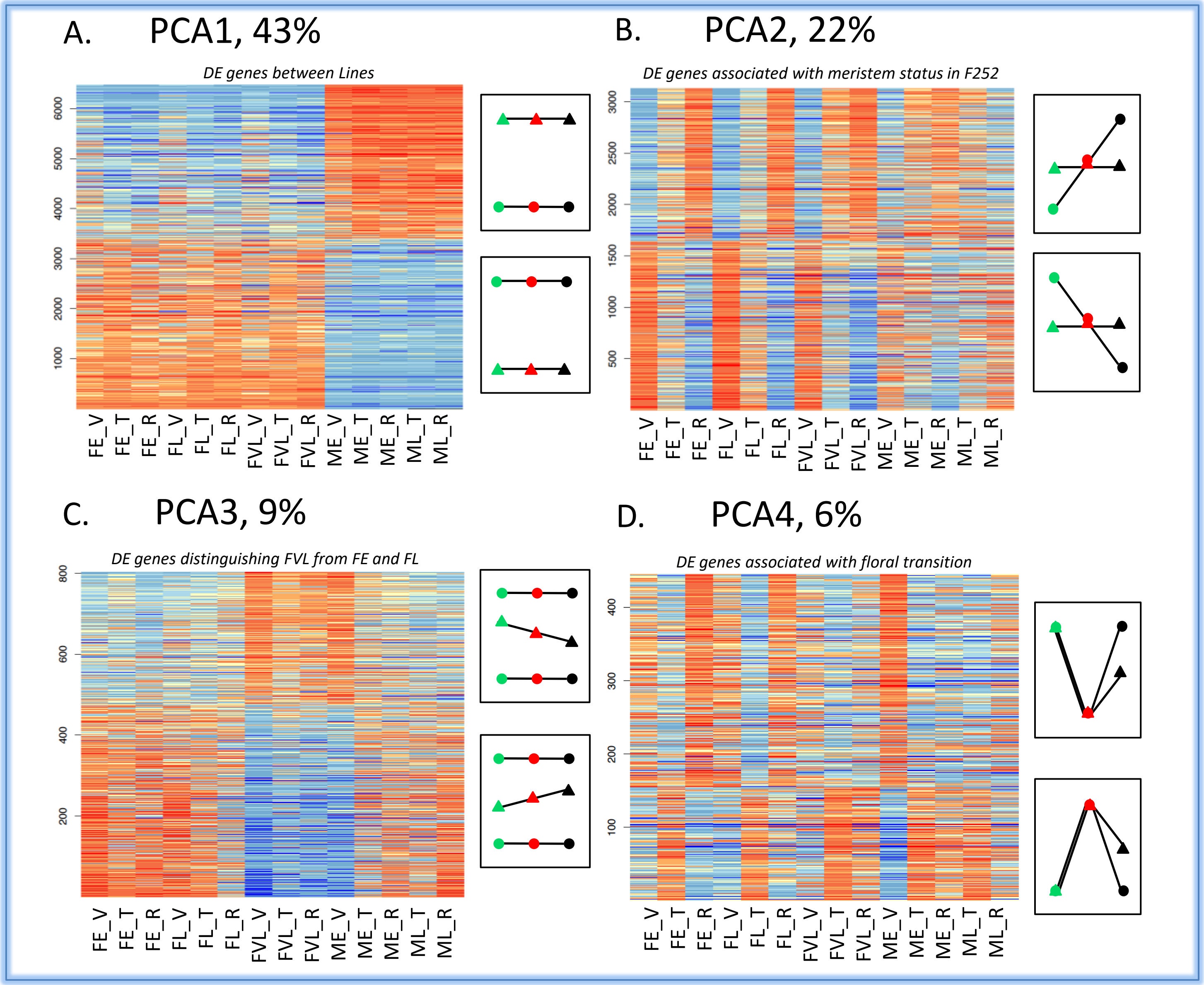
CONVERGENCE

157 candidates for flowering time
95 are DE genes (60%)
29 (18%)

ENRICHMENT

# Heat maps of Differentially Expressed (DE) genes

DE genes with greatest correlation with the 4 first PCs (normalized average expression corrected for Year)

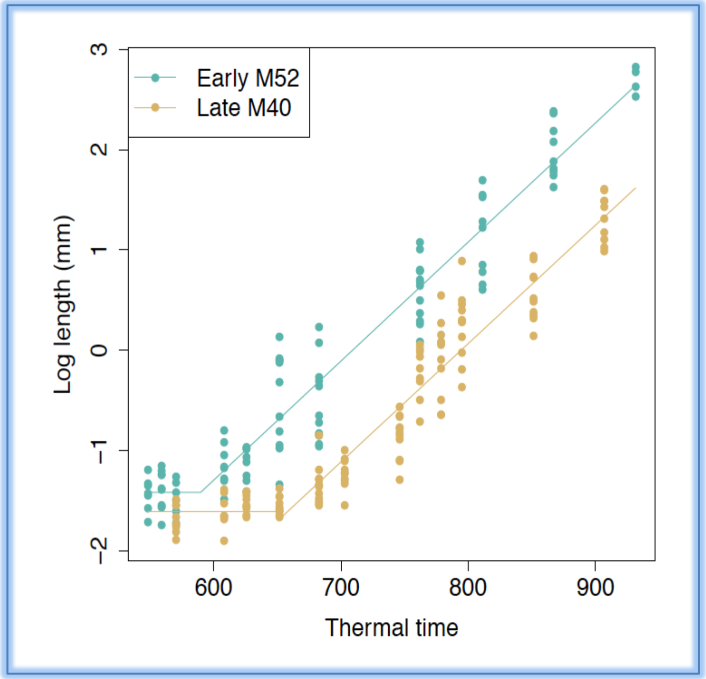


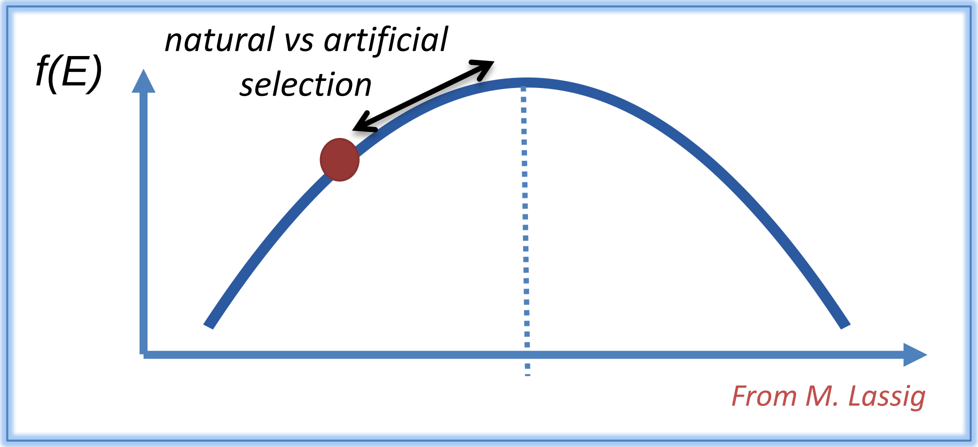
# Perspectives

Evolution of allele frequencies,
de novo vs existing variation

Integration of diverse sources of data
Transcriptomic, proteomic, development

Modeling the response to selection to understand the steadiness of the response: few polymorphisms/many DE genes, role of epistasis?





*From M. Lassig*

# Thank you to all collaborators and members of the DyGAP team

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

ANR

INVESTISSEMENTS D'AVENIR

LabEx BASC

Biodiversity, Agroecosystems, Society, Climate

UC MEXUS

ESTABLISHED 1980

Le Moulon

GÉNÉTIQUE QUANTITATIVE ET ÉVOLUTION

IDEEV

Institut Diversité, Ecologie et Evolution du Vivant