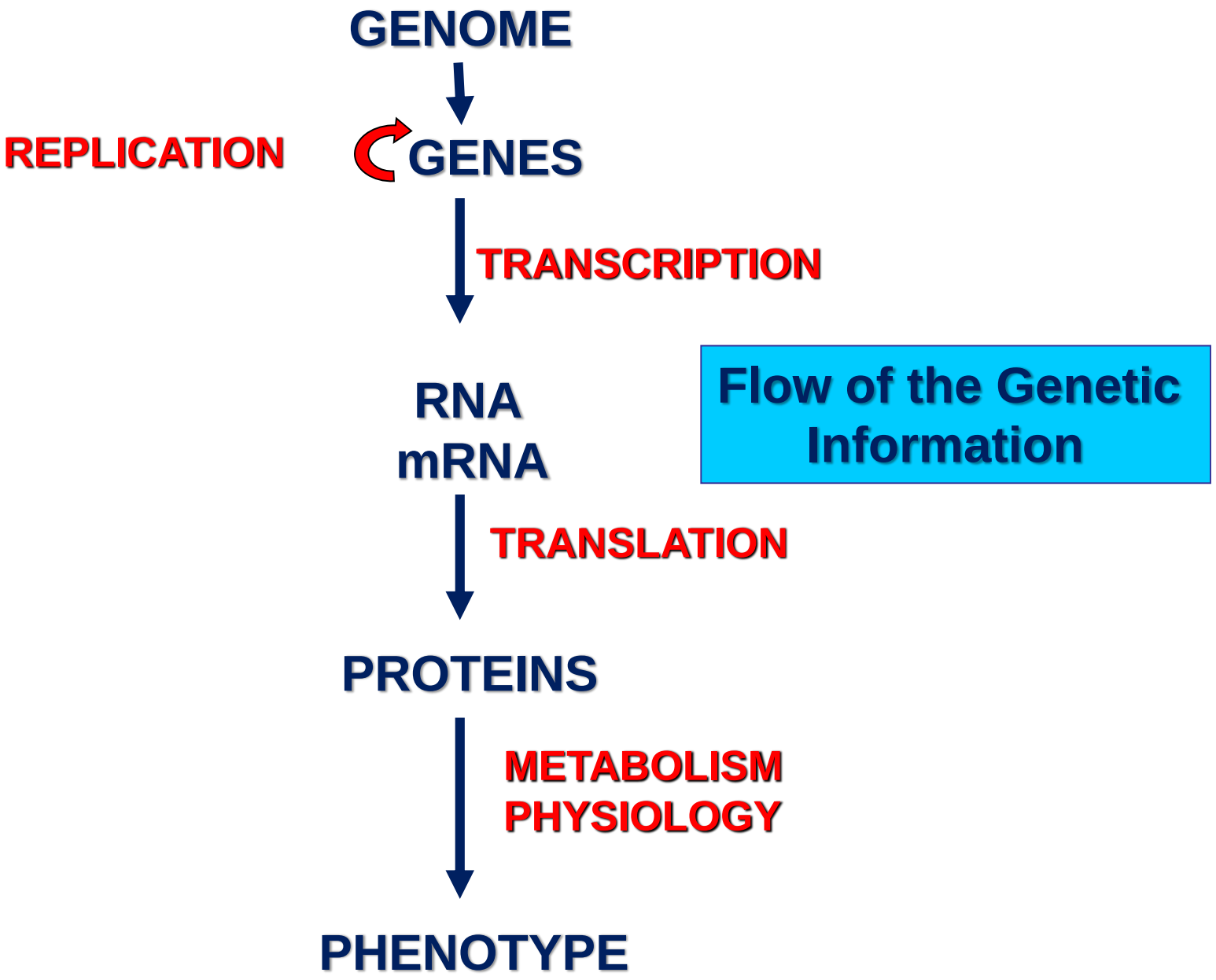
ISTITUTO
DI SCIENZE
DELLA VITA

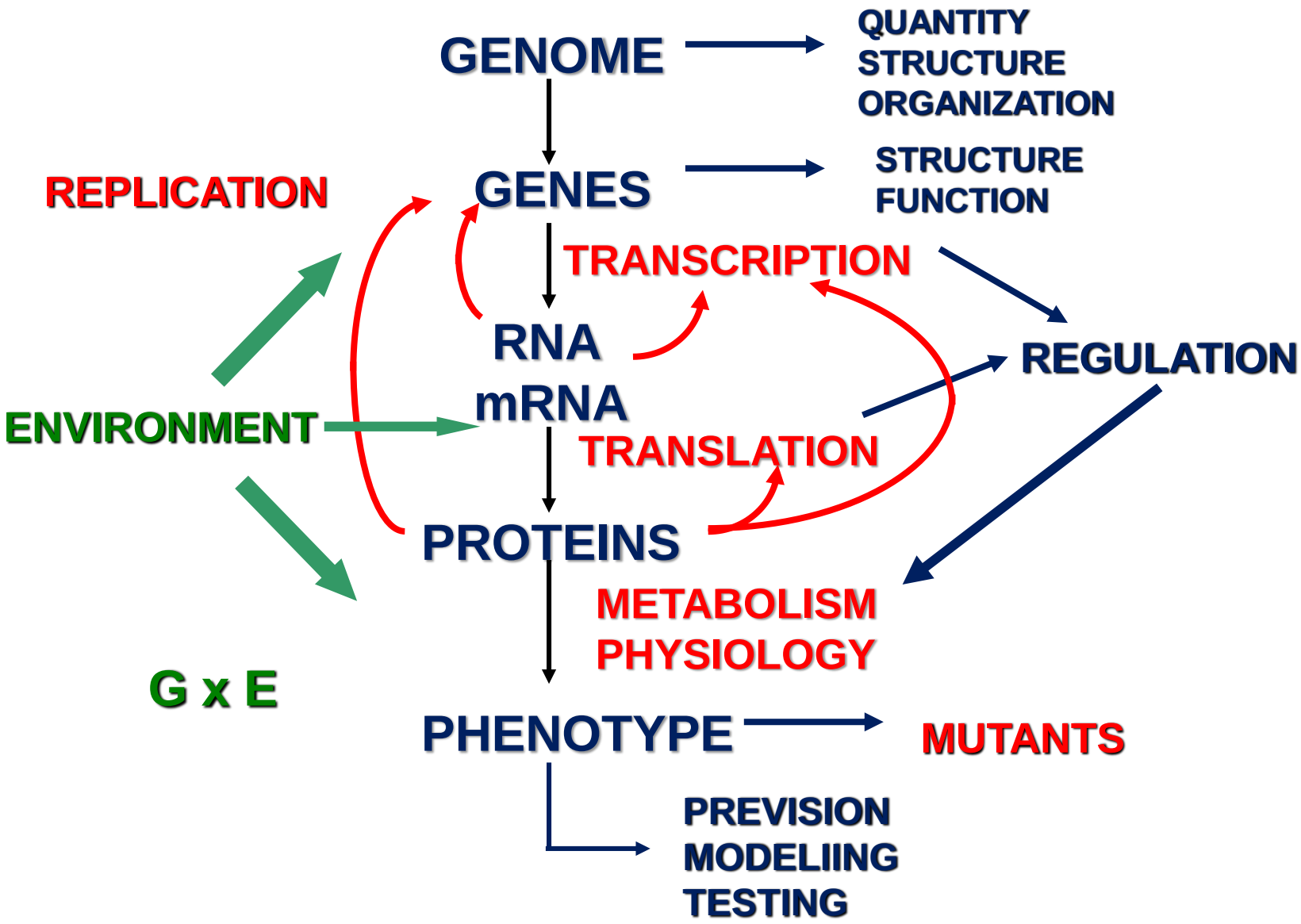
Scuola Superiore
Sant'Anna

Genotype - Phenotype: still the key question in Life Sciences

M. Enrico Pè
Scuola Superiore Sant'Anna, Pisa


GENOME
REPLICATION
GENES
TRANSCRIPTION
RNA
mRNA
Flow of the Genetic Information
TRANSLATION
PROTEINS
METABOLISM
PHYSIOLOGY
PHENOTYPE

Flow of the Genetic Information



GENOTYPE- PHENOTYPE RELATIONSHIPS: STILL THE QUESTION

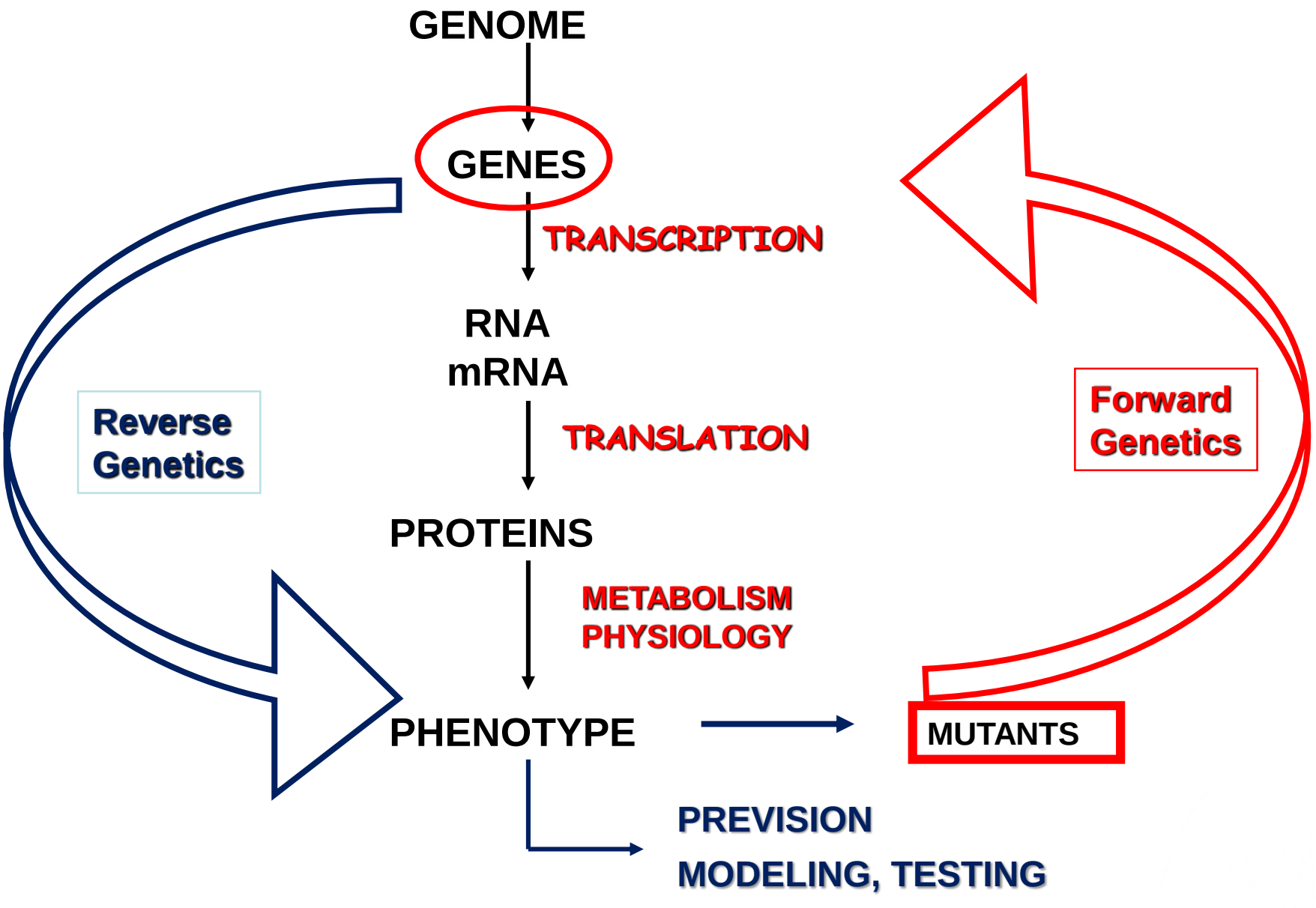
To understand the nature of genetic variation, which is at the basis of evolution

DIFFERENT SCHOOLS

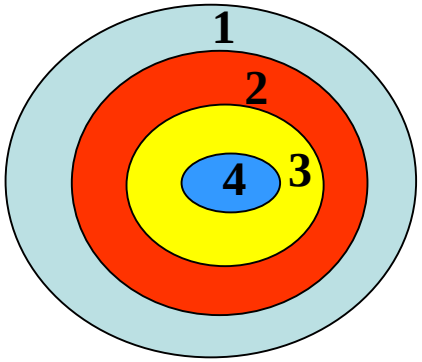
1. DARWINIAN SCHOOL
Approach: Olistic
Methods: Biometry and Quantitative Genetics

2. MENDELIAN SCHOOL
Approach: Reductionist
Methods: Mendelian Genetics, MolecularGenetics

Genetic Approaches



THE POWER OF FORWARD GENETICS: FLOWER DEVELOPMENT in *Arabidopsis thaliana*



Arabidopsis thaliana

whorl = a circular arrangement of leaves or flowers or other organs radiating from a single node

Concentric whorls

1. **Sepals**
2. **Petals**
3. **Stamens**
4. **Carpels**

The German poet Goethe in XVIII century hypothesized that a flower is a transformed leaf

Single-gene Homeotic Mutations Disrupt the Genetic Regulation of Organ Identity

Mutant *apetala 1*

Mutant *pistillata*

Mutant *agamous*

ABC Model

nature
International journal of science

Altmetric: 11 Citations: 1728

Review Article | Published: 05 September 1991

The war of the whorls: genetic interactions controlling flower development

Nature 353, 1991

Enrico S. Coen & Elliot M. Meyerowitz

Link Between Forward Genetics and Reverse Genetics

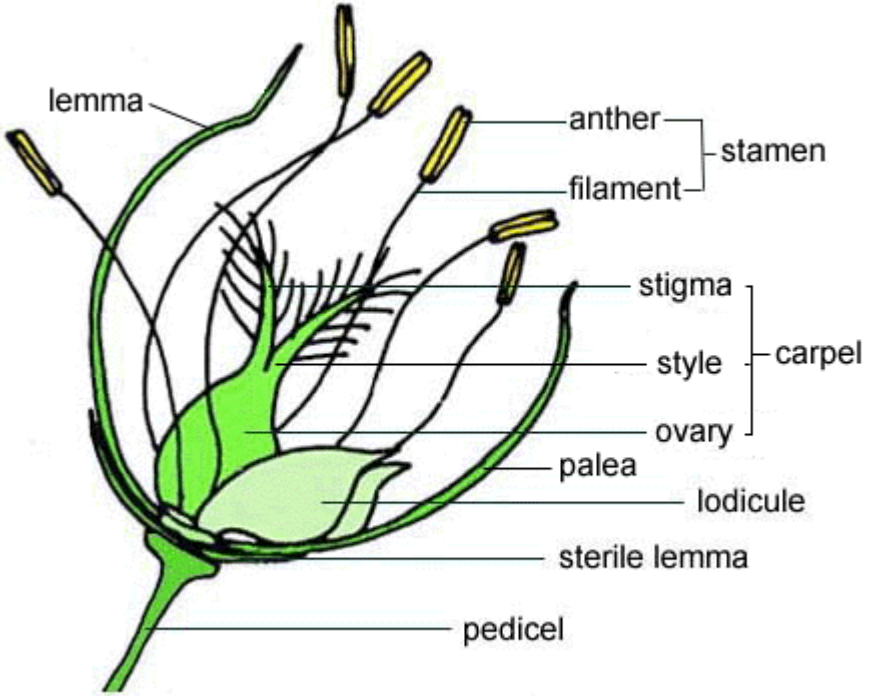
Homeotic genes code for transcription factors controlling flower development

Most of homeotic genes belong to the MADS box gene family

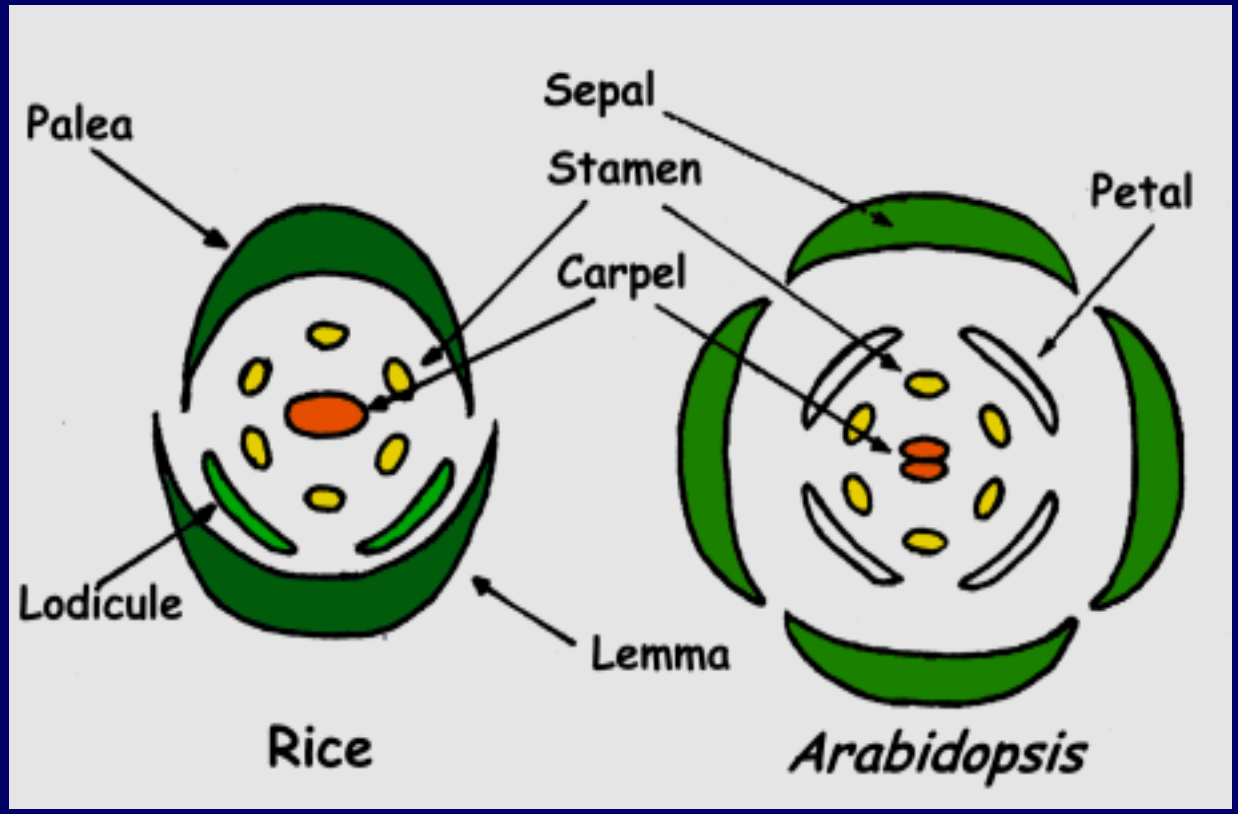
MADS is a functional domain of about 60 AA for binding to DNA HIGHLY CONSERVED

Starting from this information it is possible to isolate MADS genes from other species and test the model in other species as well

Arabidopsis



Rice spikelet


Palea
Sepal
Stamen
Petal
Carpel
Lodicule
Lemma
Rice
Arabidopsis

Quantitative Traits

The most important for breeding and evolution

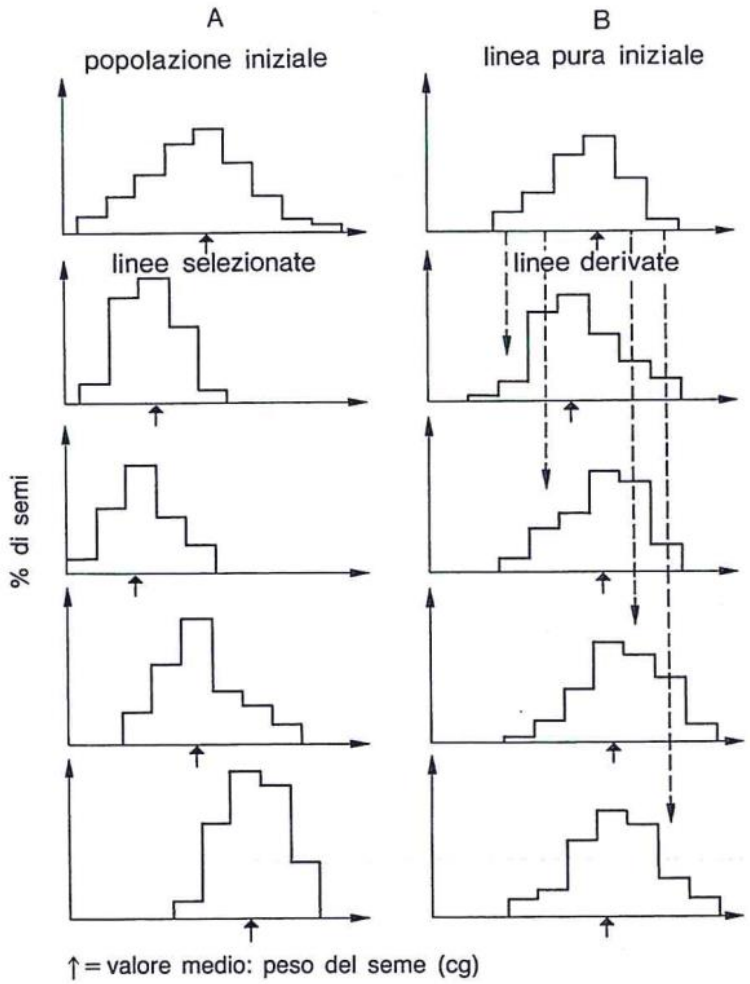
Mendelian Analysis cannot be applied

Differences among individuals are measured

Phenotypic values depend on genetic and environmental factors

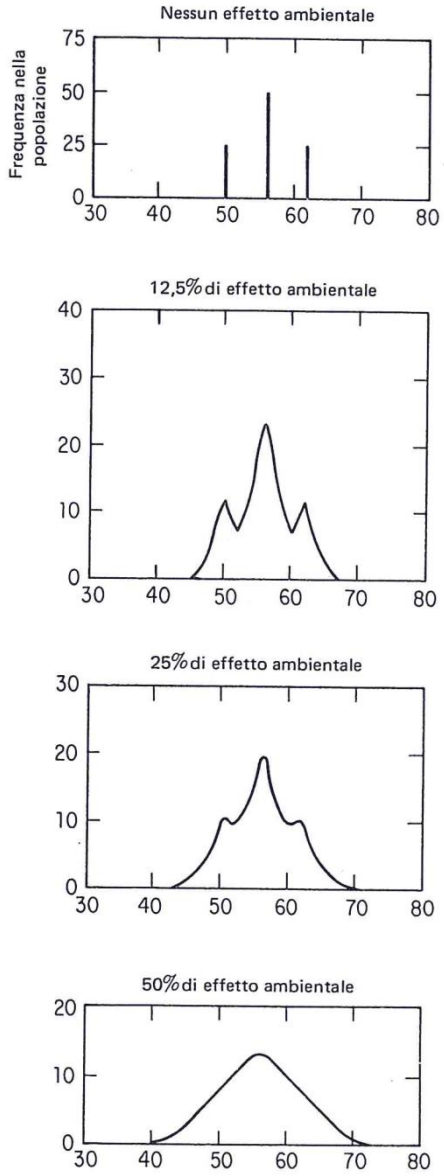
Genetics and Environment

Wilhelm Johannsen (1857-1927)
Danish botanist
Coined the terms Gene, Genotype and Phenotype



1903: the beans pure lines experiment

The Confounding Effect of the Environment



With strong environmental effects even the action of a single segregating gene can be blurred

Theory of Polygenic Inheritance of Quantitative Traits and the Application of Biometry

Sir Ronald Fisher
(1890-1962)
British statistician
and geneticist
Great contribution to
Population genetics
and New Darvinism

1. Phenotypic value depends on G and E
2. Genetic effects are due to the action of many genes with additive effects
3. Effect of a single gene is small
4. Environment has a stronger effect than that of a single gene

CONSEQUENCES

1. P = G + E
2. G is an estimate of all genetic effects
3. In order to estimate G, E must also be estimated
4. Application of biometrical analysis MOSTLY BASED on the analysis of variance

Genetic Meaning of Variance

P = phenotypic value

G = genotypic value,

E = environmental effects

P = G + E, if E acts randomly on the different individuals

Considering many individuals:

Phenotypic variation : $\sigma^2_P = \sigma^2_G + \sigma^2_E$

$\sigma^2_G = 0$ if individuals are genetically uniform

Hereditability in broad sense $h^2_b = \sigma^2_G / \sigma^2_P$

Dissection of the Genetic Variance

σ^2_G can be used to predict the effect of selection (natural and artificial) only in genetically homogeneous and strictly autogamous species, if this is not the case:

- Individuals transmit thier genes not their genotype
- Segregation, recombination and epistasis contribute in determining parents/offspring resemblance
- Model to partition σ^2_G in different components

$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$ where

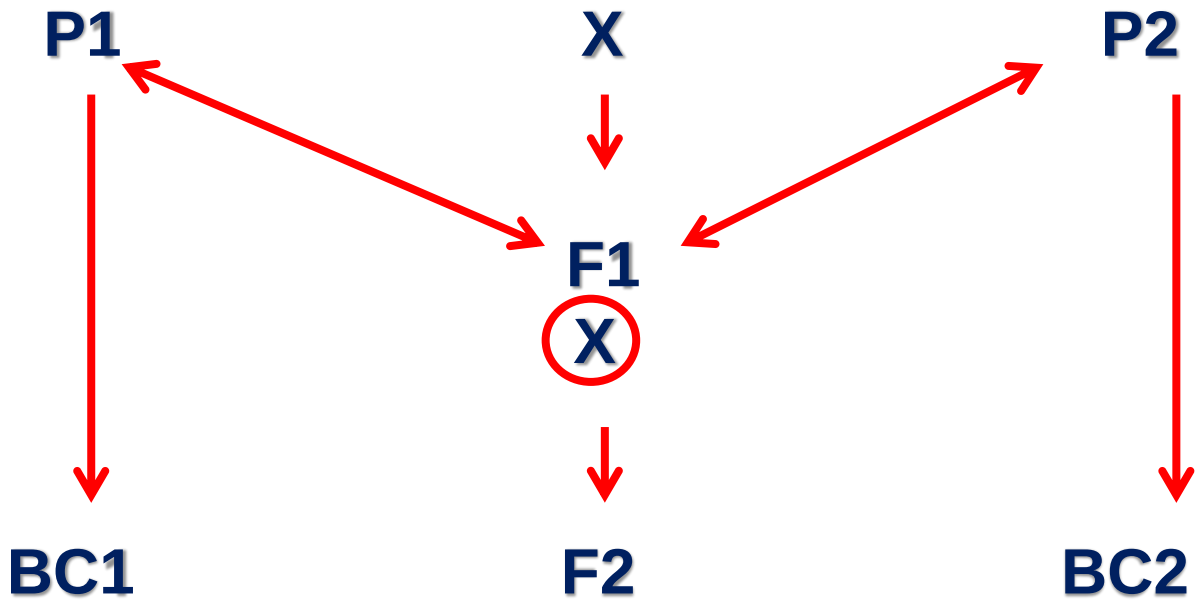
σ^2_A = additive variance

σ^2_D = dominance variance (interactions within *loci*)

σ^2_I = variance due to interactions among different *loci*

Mating Design

Populations with different genetic structure are produced



Different populations allow estimating σ^2_G components

Inbred lines: σ^2_A; among F1: σ^2_D

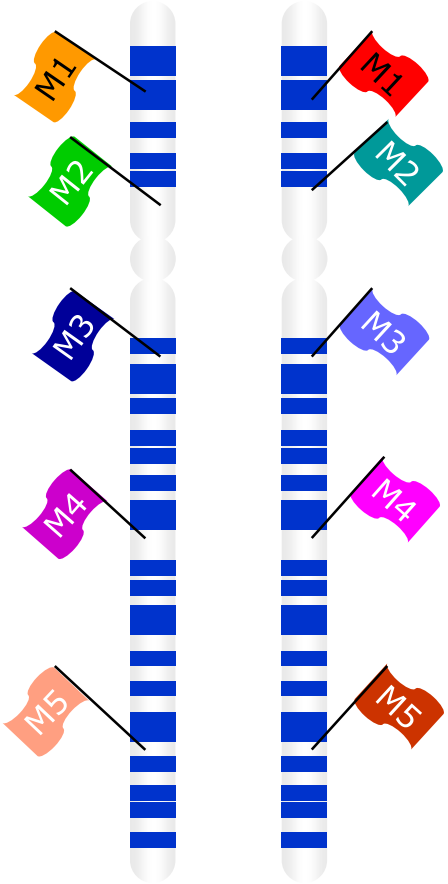
F2: : $\sigma^2_A + \sigma^2_D$; BC1 + BC2: $\sigma^2_A + 2\,\sigma^2_D$

Quantitative Genetics

- **Mating design: comparison among artificial population with different genetic structure**
- **Experimental design: accurate evaluation of phenotypic values**

- **Estimate of genetic effects vs environmental effects**
- **Information on the genetic system as a whole**
- **No information on single genes effect or gene number or gene position**

The Revoution of QTL Mapping



Genetic Marker

A *Locus* which identifies un- ambigously a specific chromosomla region

- **Tool** → not the objective
- **Marker** → indirect approach

Genetic Variation and Genotyping

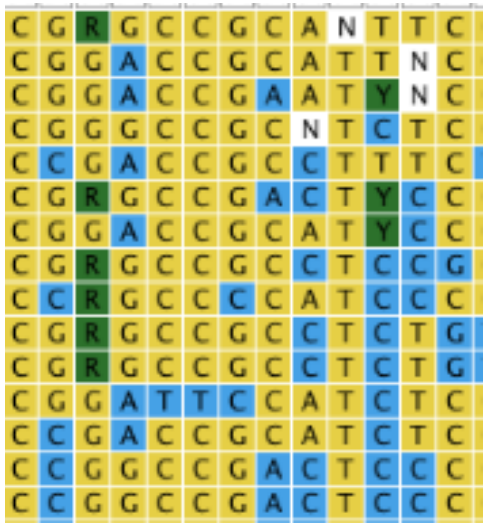
…C C **A** T T G A C…
…G G **T** A A C T G…

Allele 1

…C C **G** T T G A C…
…G G **C** A A C T G…

Allele 2

Different individuals



Genetic variants

Genetic Mapping in the Genomic Era

- **Efficient way to MAP phenotypes**
- **It contributes to better understand the relationship between genotype and phenotype**
- **Indication on the function of DNA sequences (genes; regulatory elements)**
- **Based on the linkage between DNA polymorphisms and the causal factor determining phenotypic variation:**
  - **STATISTICAL EVIDENCE**
  - **INDIRECT APPROACH**

If the phenotype is quantitative → QTL Mapping

QuantitativeTrait Loci Analysis

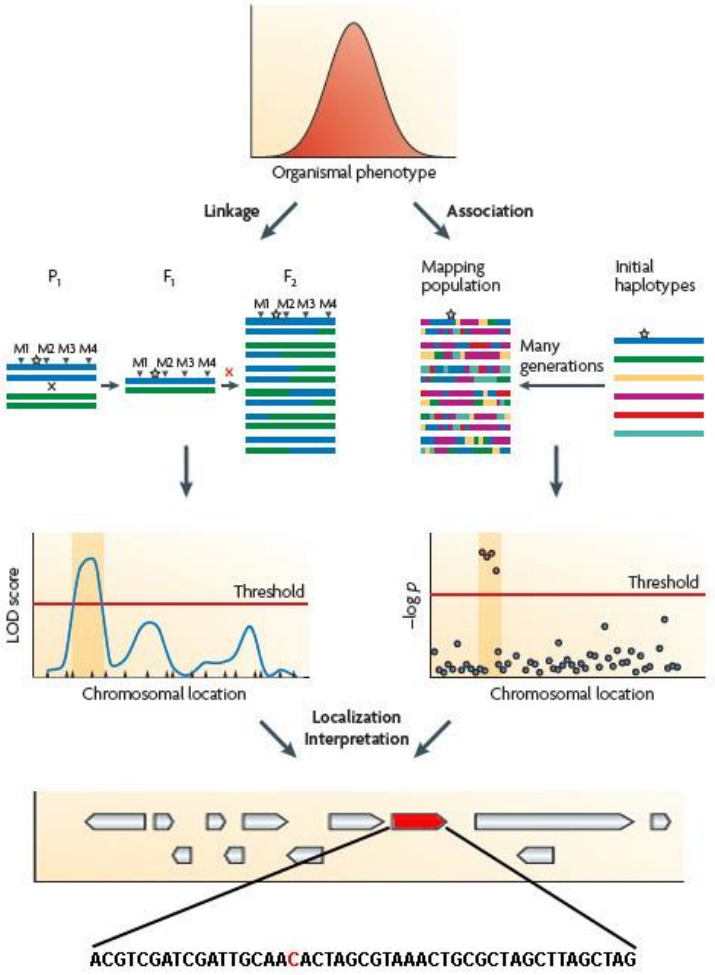
- A QTL is a *locus* contributing to the phenotypic value of a complex (multigenic) trait.
- QTL analysis aims at the dissection of complex traits into Mendelian factors.
- Advanced statistical methods are used to test whether a QTL is likely to be present at certain *loci*.
- Appropriate materials (crossing or natural populations)
- Molecular characterization of the population
- Consistent and reproducible phenotypic data

Two Different Methods to Map QTL

Linkage mapping

- low density required
- fully known pedigree
- more robust
- limited variation
- low definition
- time demanding

Structured pops

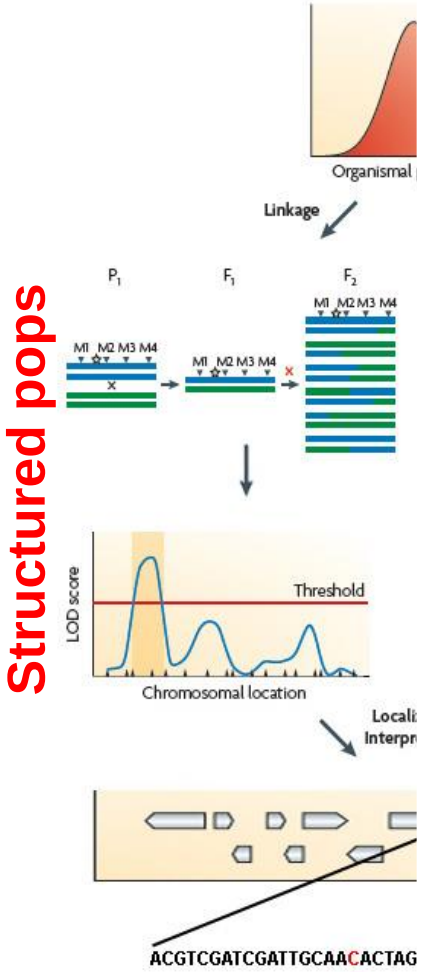


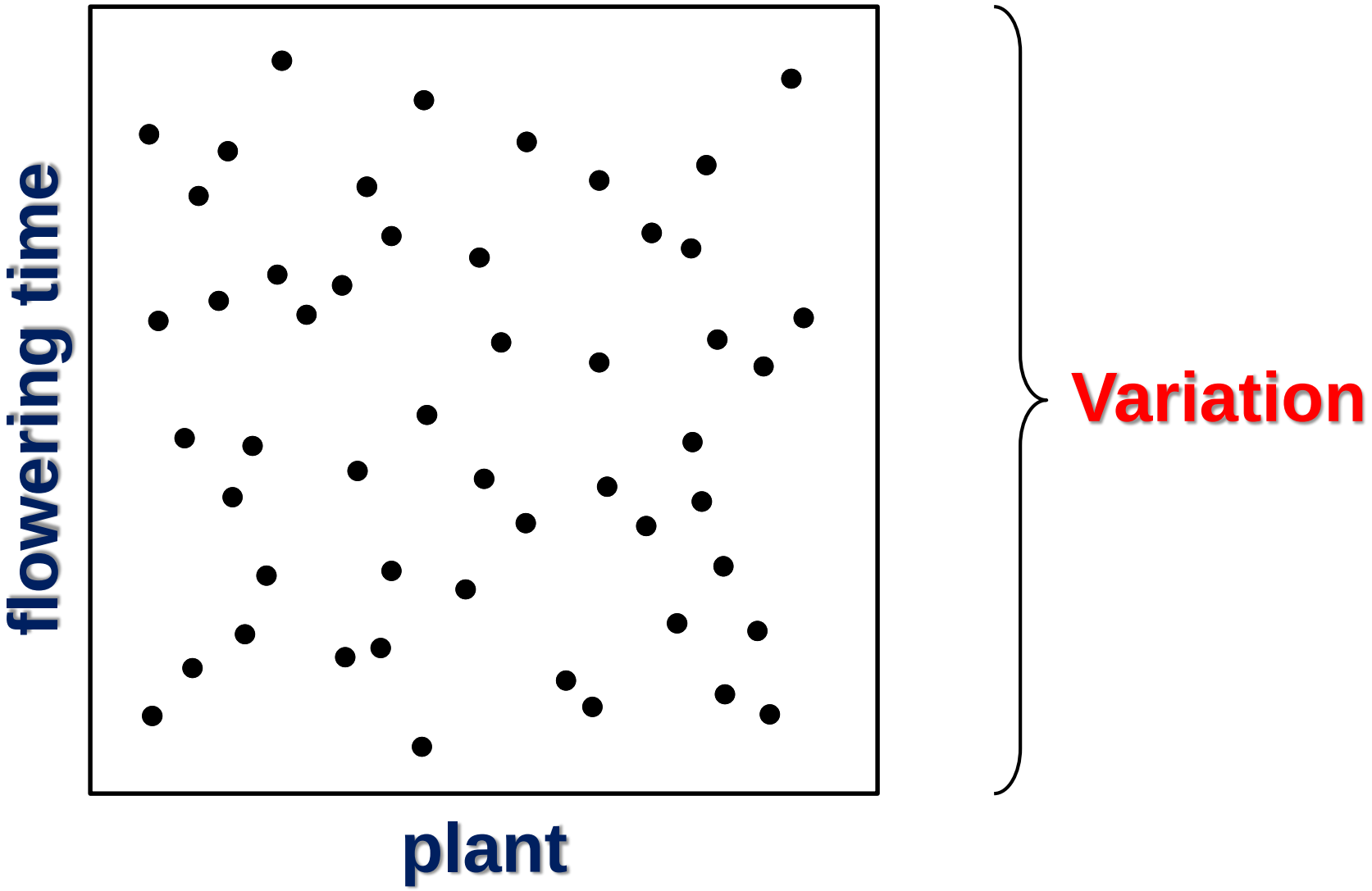
Non-structured pops

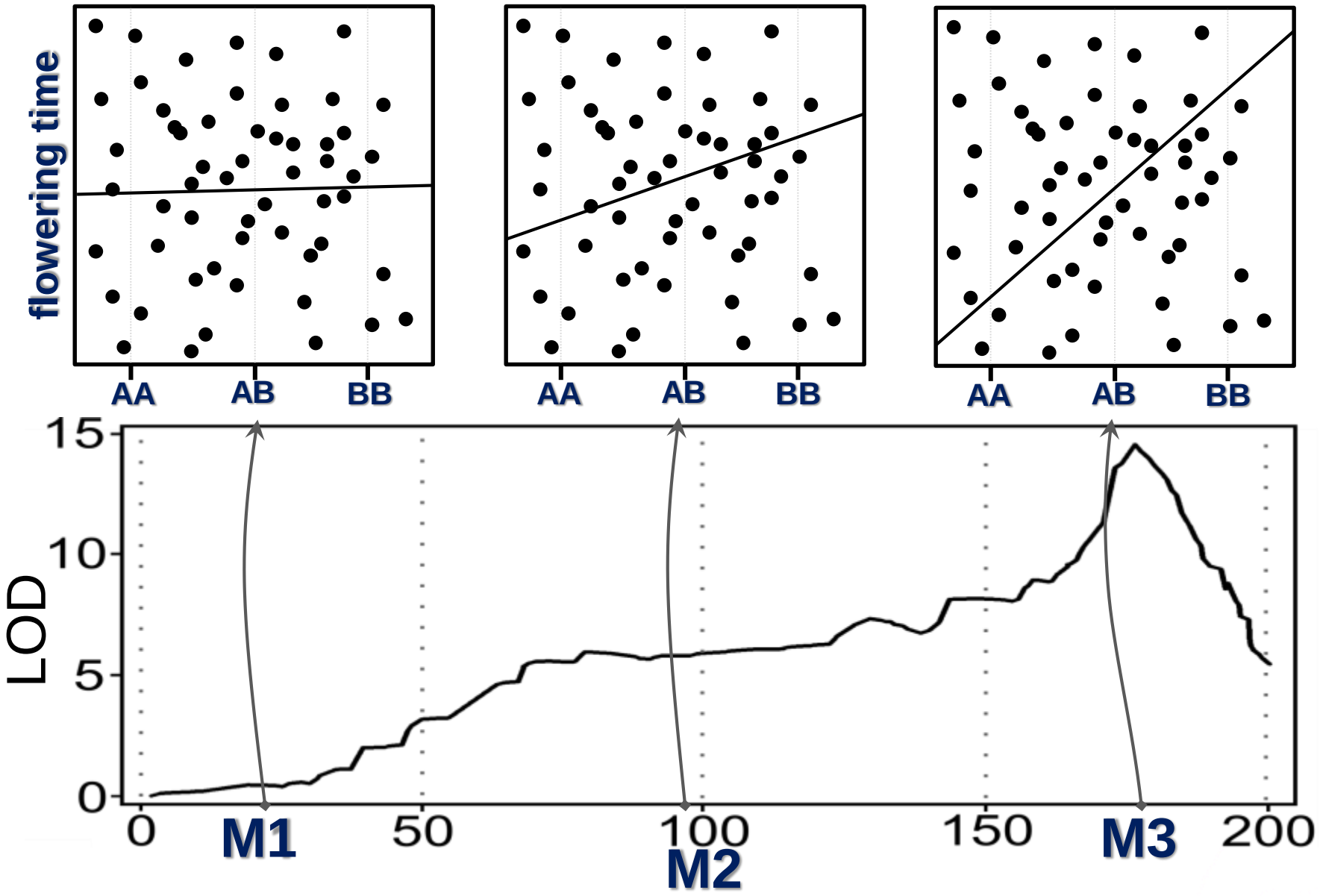
Association (GWAS)

- high density necessary
- hidden structure, LD
- higher false rate
- broad variation
- high definition
- faster, cheaper

A closer look to QTL mapping in Structured Populations




flowering time
plant
Variation

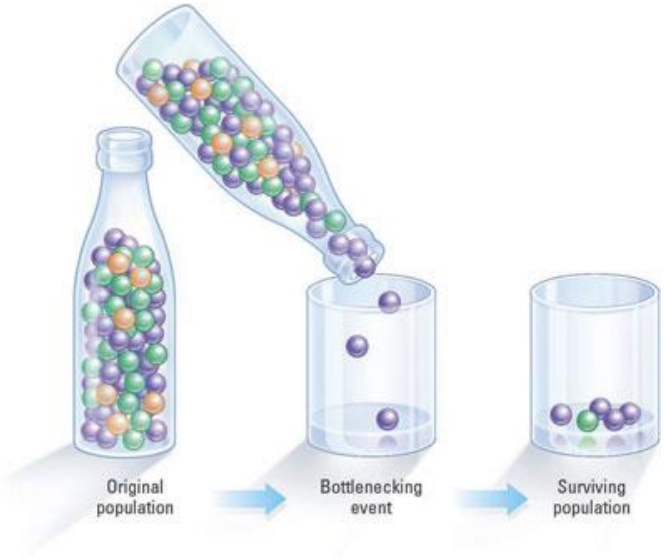

flowering time
AA
AB
BB
AA
AB
BB
AA
AB
BB
LOD
15
10
5
0
0
50
100
150
200
M1
M2
M3

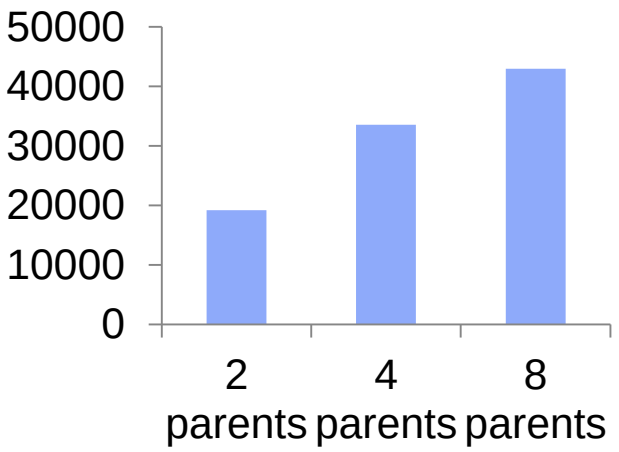
Two major features are required for QTL mapping to be efficient:

1. The amount of genetic/phenotypic diversity (amount of QTL detected)
2. The amount of recombinations (definition by which QTL are detected)

1

More diversity means more alleles into play (more segregating traits)

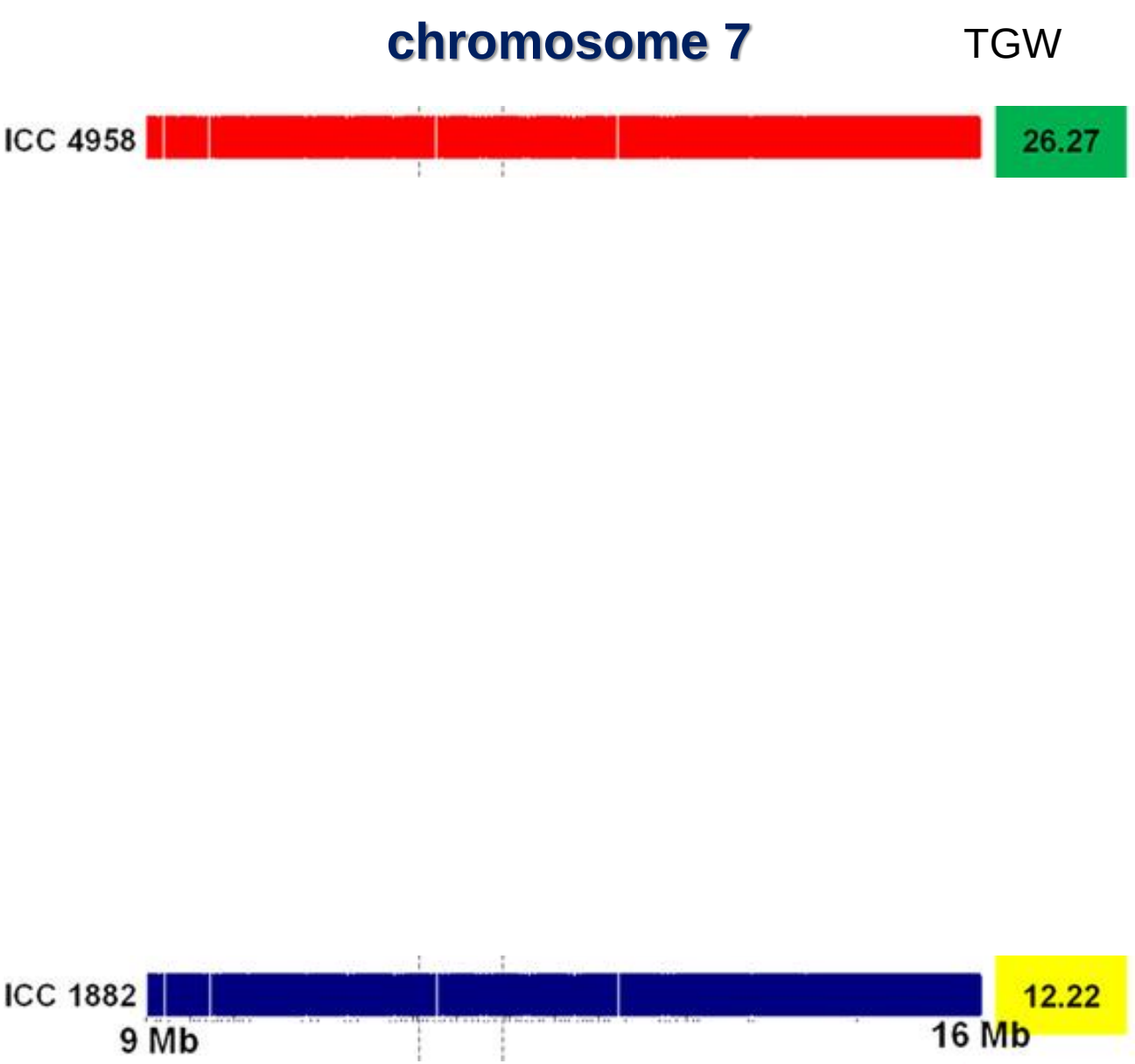




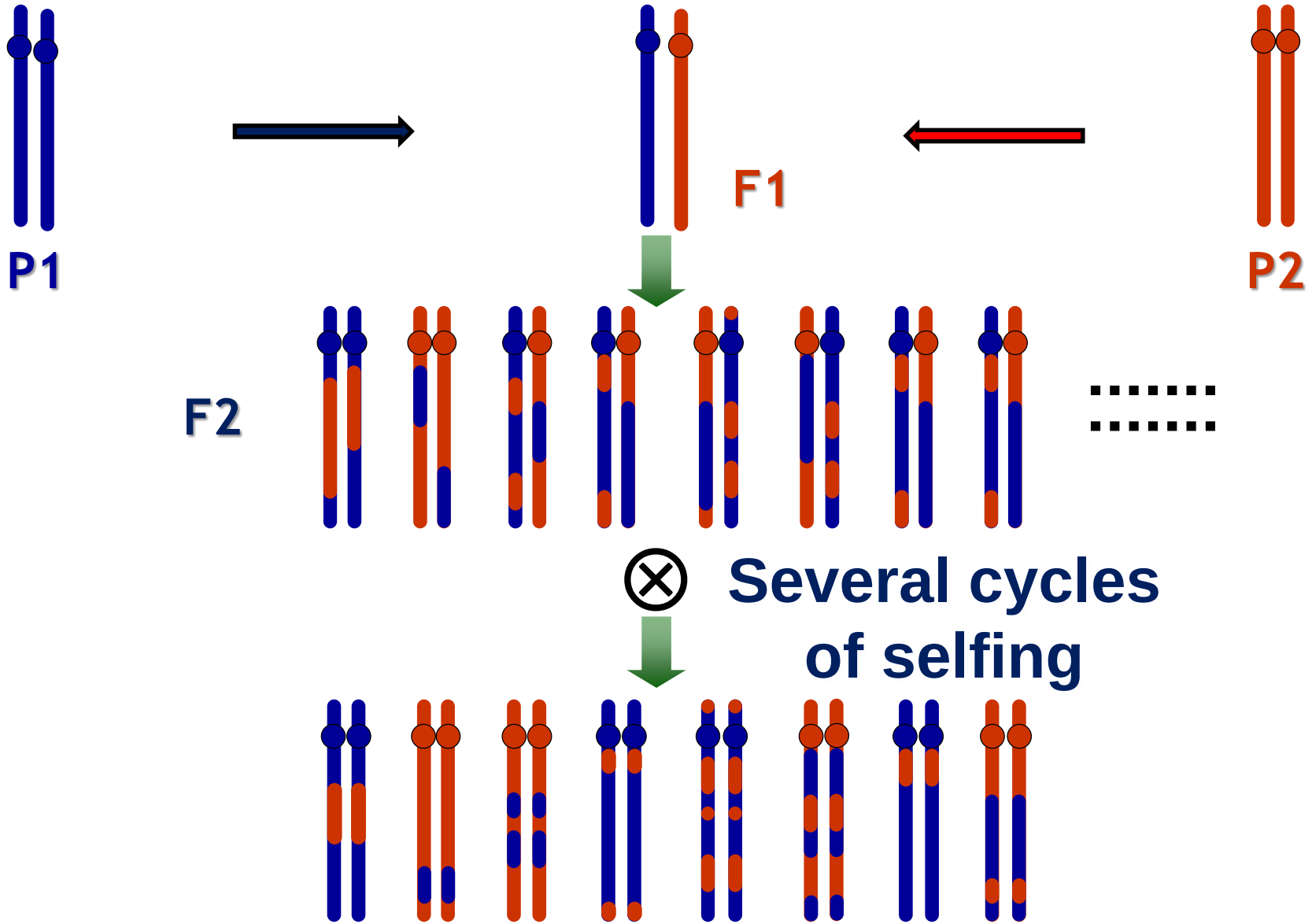
2

Imagine a QTL for Yield somewhere on this chr. Red parent high, blue parent low. How to pinpoint it?

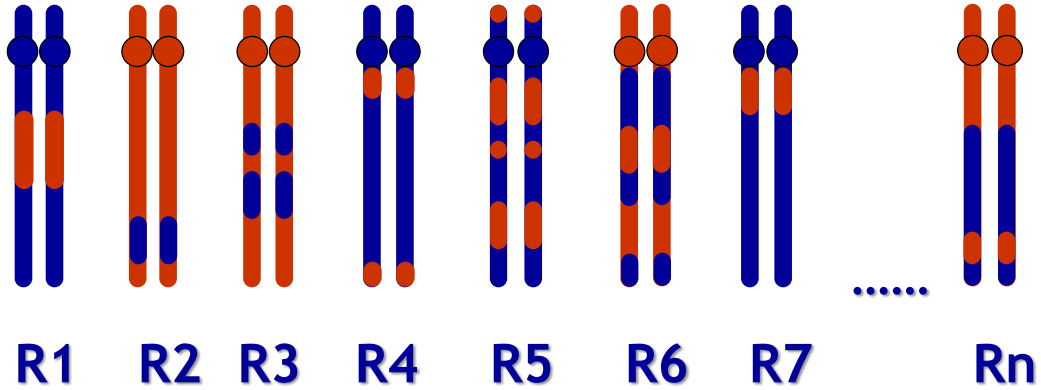
the rationale of having a segregant population is to break founder genomes in smaller recombination blocks



Recombinant Inbred Lines (RIL)



Recombinant Inbred Lines (RIL)



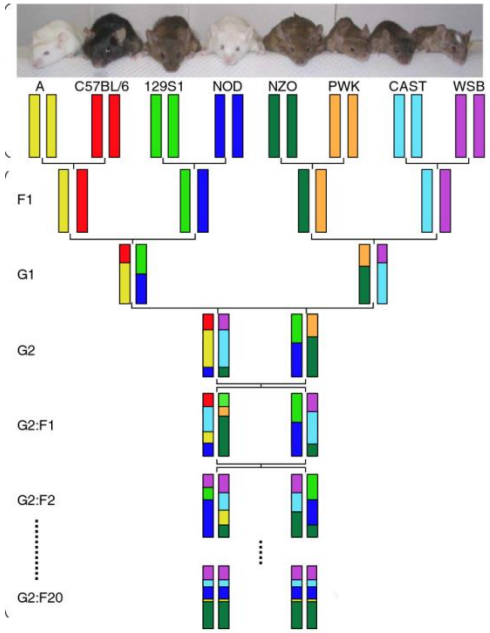
Immortalized genotypes: recombination blocks are fixed

RILs are poor in both

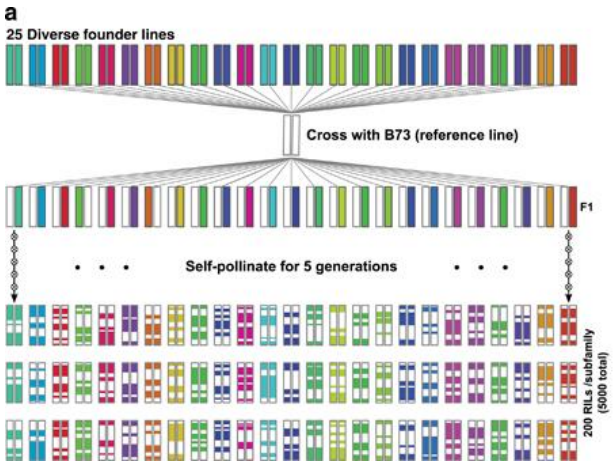
- Variability

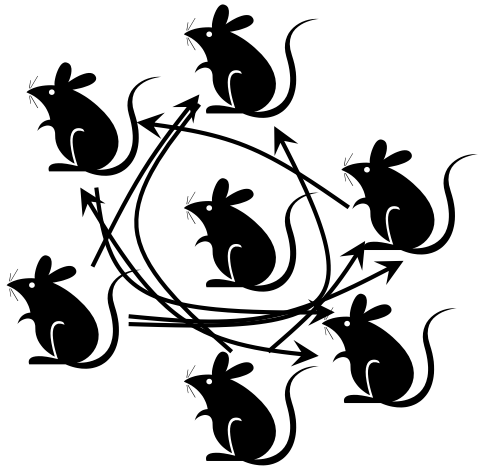
- N°of informative recombinations

Multiparental Populations Achieve More Diversity and More Recombination Events than Biparental populations



…with different designs





Collaborative Cross-like

NAM-like

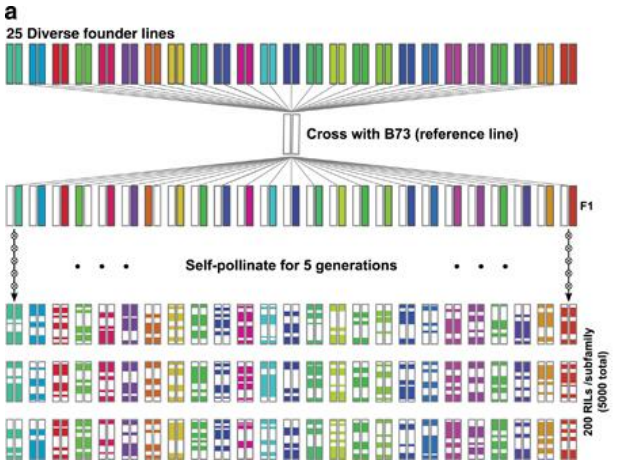
Outbred-like

and variation on the theme

The Genetic Architecture of Maize Flowering Time

Edward S. Buckler,[1,2,3*] James B. Holland,[1,4*] Peter J. Bradbury,[1,2] Charlotte B. Acharya,[2] Patrick J. Brown,[2] Chris Browne,[1,5] Elhan Ersoz,[2] Sherry Flint-Garcia,[1,5] Arturo Garcia,[1,5] Jeffrey C. Glaubitz,[6] Major M. Goodman,[4] Carlos Harjes,[7] Kate Guill,[1,5] Dallas E. Kroon,[2] Sara Larsson,[3] Nicholas K. Lepak,[1,3] Huihui Li,[8,2,9] Sharon E. Mitchell,[2] Gael Pressoir,[2] Jason A. Peiffer,[3] Marco Oropeza Rosas,[4] Torbert R. Rocheford,[10,11] M. Cinta Romay,[2,12] Susan Romero,[2] Stella Salvo,[1,4] Hector Sanchez Villeda,[5,13] H. Sofia da Silva,[10] Qi Sun,[14] Feng Tian,[2] Narasimham Upadyayula,[10] Doreen Ware,[1,15] Heather Yates,[2] Jianming Yu,[16] Zhiwu Zhang,[2] Stephen Kresovich,[2*] Michael D. McMullen[1,5*]

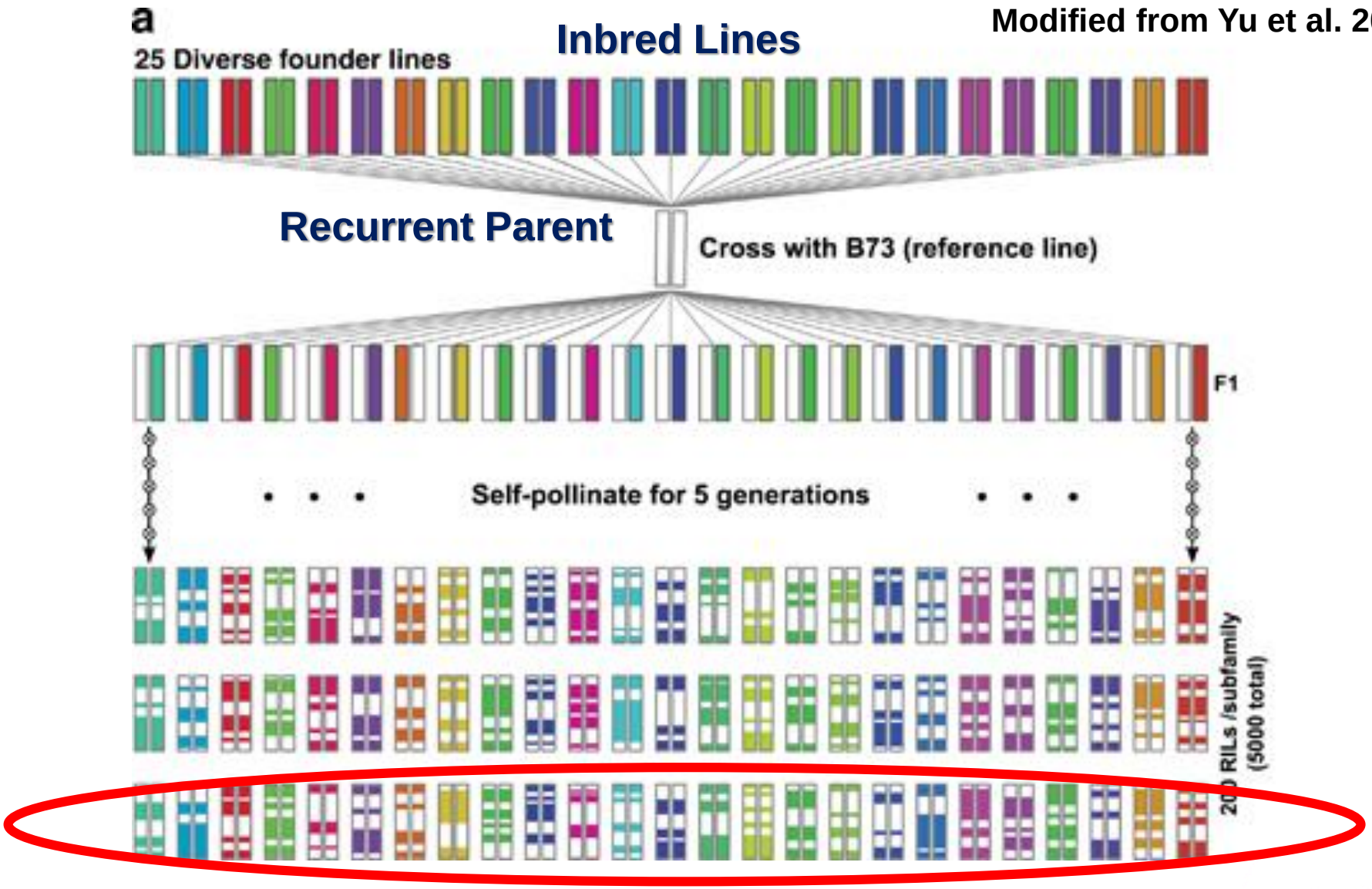
7 AUGUST 2009 VOL 325 **SCIENCE**



Nested Association Mapping - NAM - Population

Nested Association Mapping
- NAM - Population

Modified from Yu et al. 2008



NAM POPULATION CHARACTERISTICS

Advantages

- Easy to produce
- A lot of variation
- Open scheme

Disadvantages

- Genetic variation dispersed
- It does NOT increase the number of recombinations / RIL
- It requires high number of RIL to achieve high definition

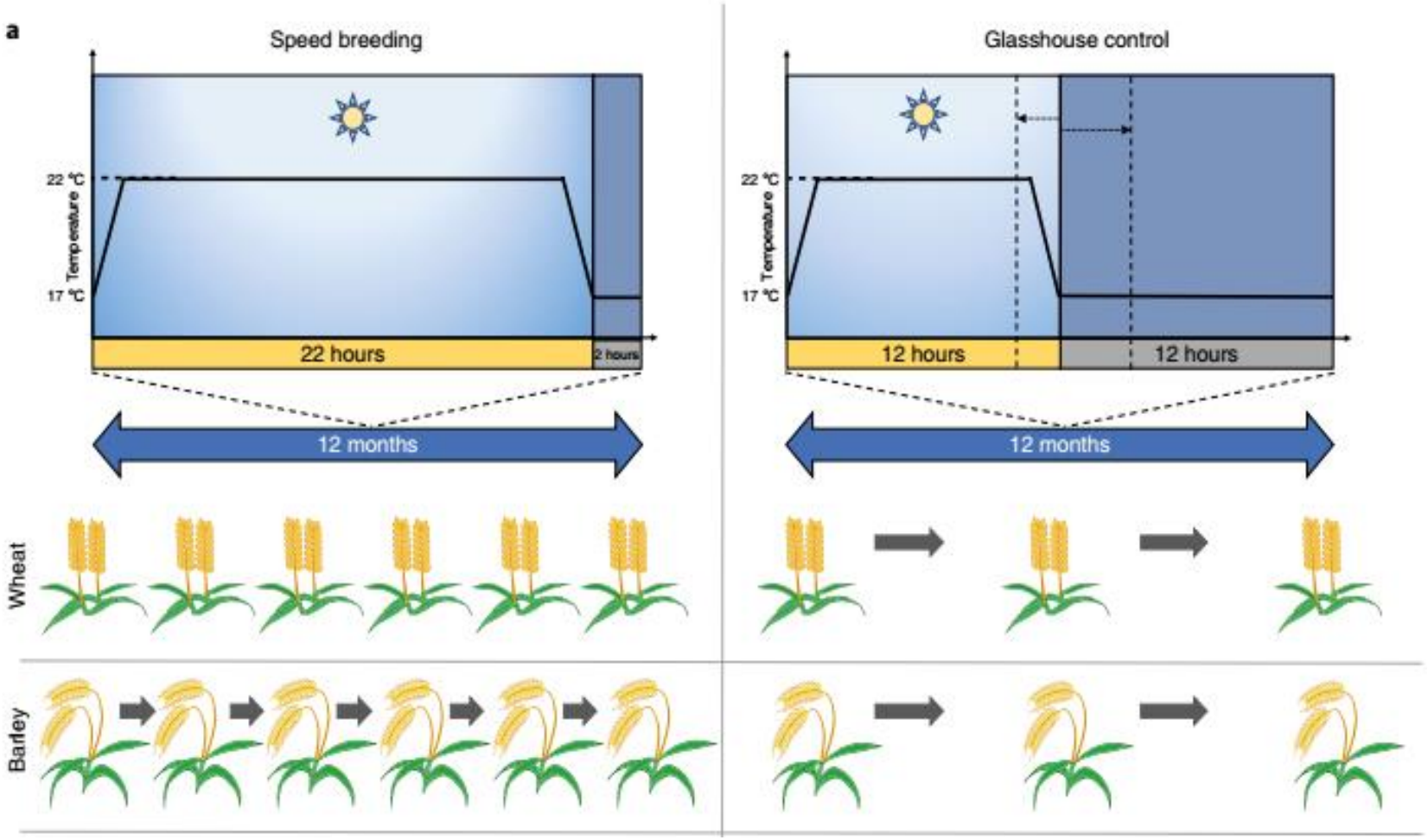
NAM Populations Currently Publicly Available

Species	Founders	# of Lines	Notes	Reference
Barley	25 wild x elite	1,420	BC-NAM 5,700 SNP	Maurer et al. 2016
Spring wheat	9 x elite	852	Stem rust	Bajgain et al. 2016
Sorghum	10 x recurrent	2,310	6,000 SNP chip	Bouchet et al. 2017
Soybean	40 x elite	5,600	90,000 SNP chip	Song et al . 2017
Durum wheat	50 landrace x elite	6,280	1,200 9,000 SNP chip	Kidane et al. submitted

nature plants

LETTERS

https://doi.org/10.1038/s41477-017-0083-8



Dr. Lee Hickey - University of Queensland

ETHIOPIA: Valorization of durum wheat genetic resources

- 100? Million people (2°in Africa)
- 83% of the population lives of subsistence farming in rural areas (accounting for 47% of GDP)
- Per capita GDP is 550 $
- 40% of the population is undernourished

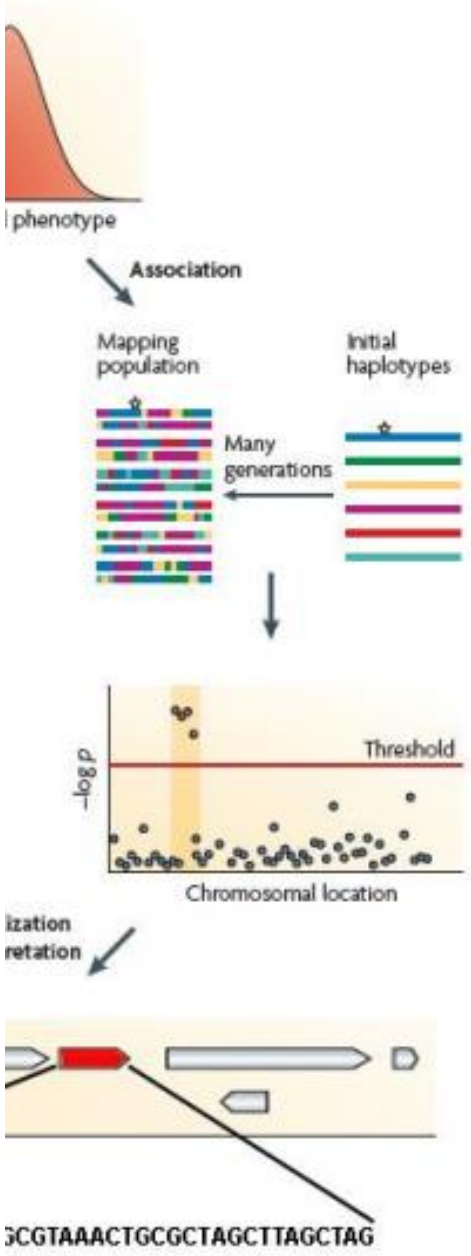
Ethiopia is a Vavilov center for crop diversity

Coffee, barley, millet, sesame, teff, **durum wheat**

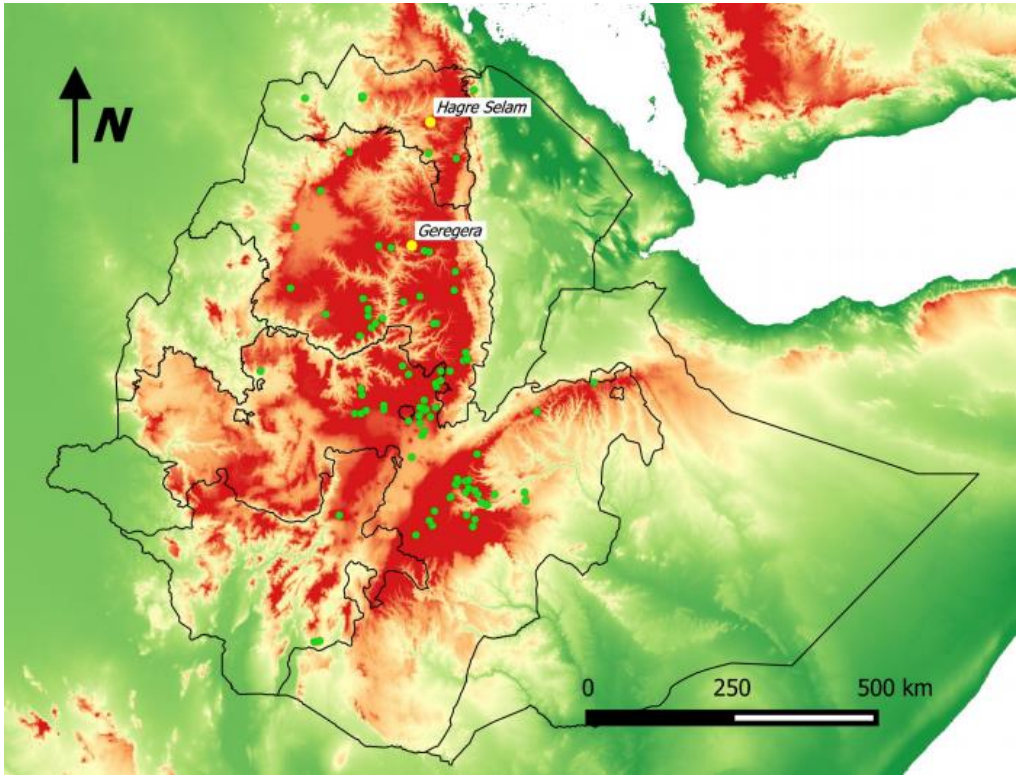
Diversity is mostly mantained in thousands of landraces cultivated within the family farming system, partially mantained at the Ethiopian Biodiversity Institute

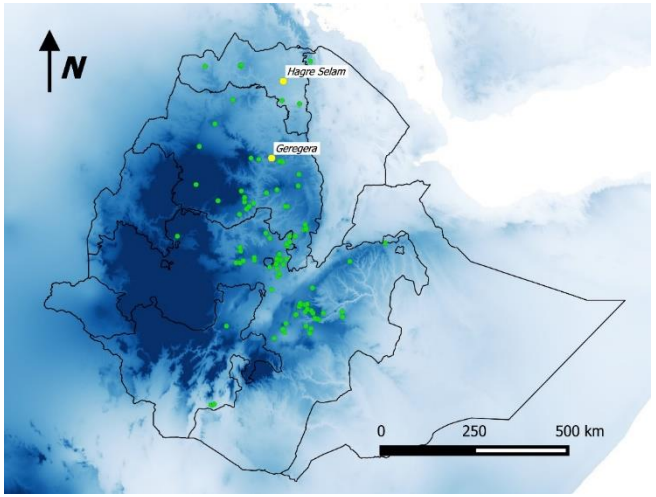
Diversity Panels

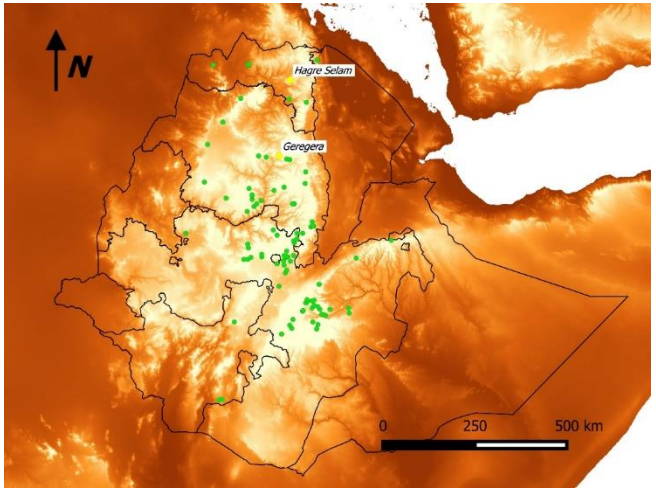
Non-structured pops



The massive diversity in Ethiopian landraces is linked to the varying environmental conditions in which durum wheat is cultivated in rainfed, low input fields → landraces are adapted to local climate

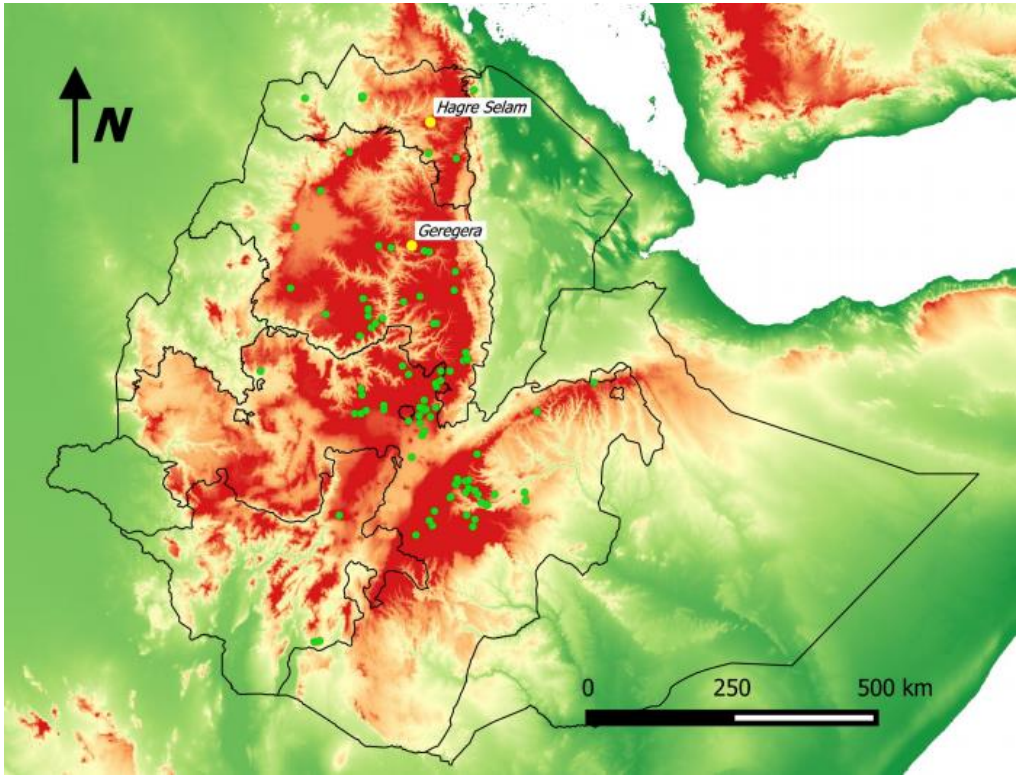




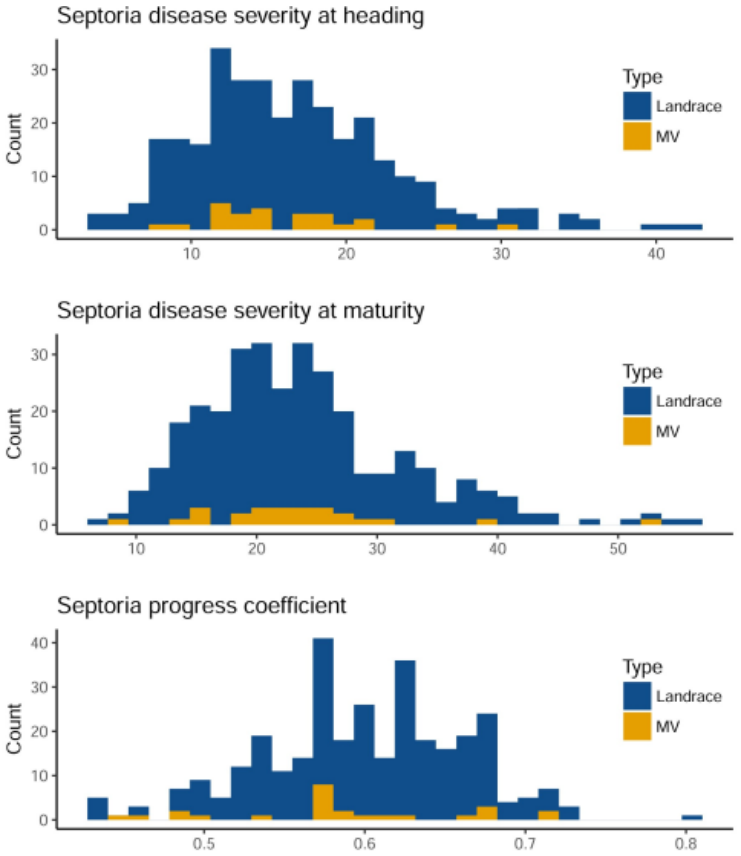


Landscape genomics approaches can be used to track the molecular basis of environmental adaptation of landraces

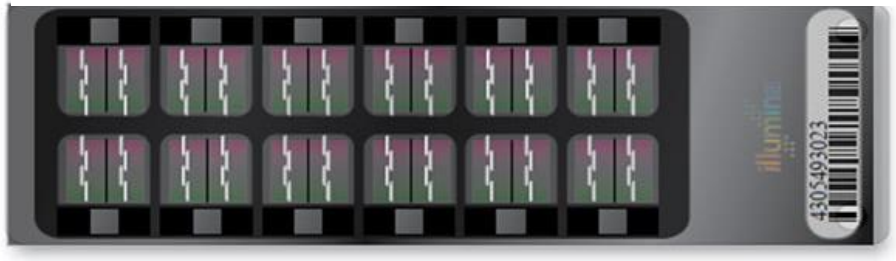
ASSEMBLY and CHARACTERIZATION of a DIVERITY PANEL for GWAS ANALYSIS



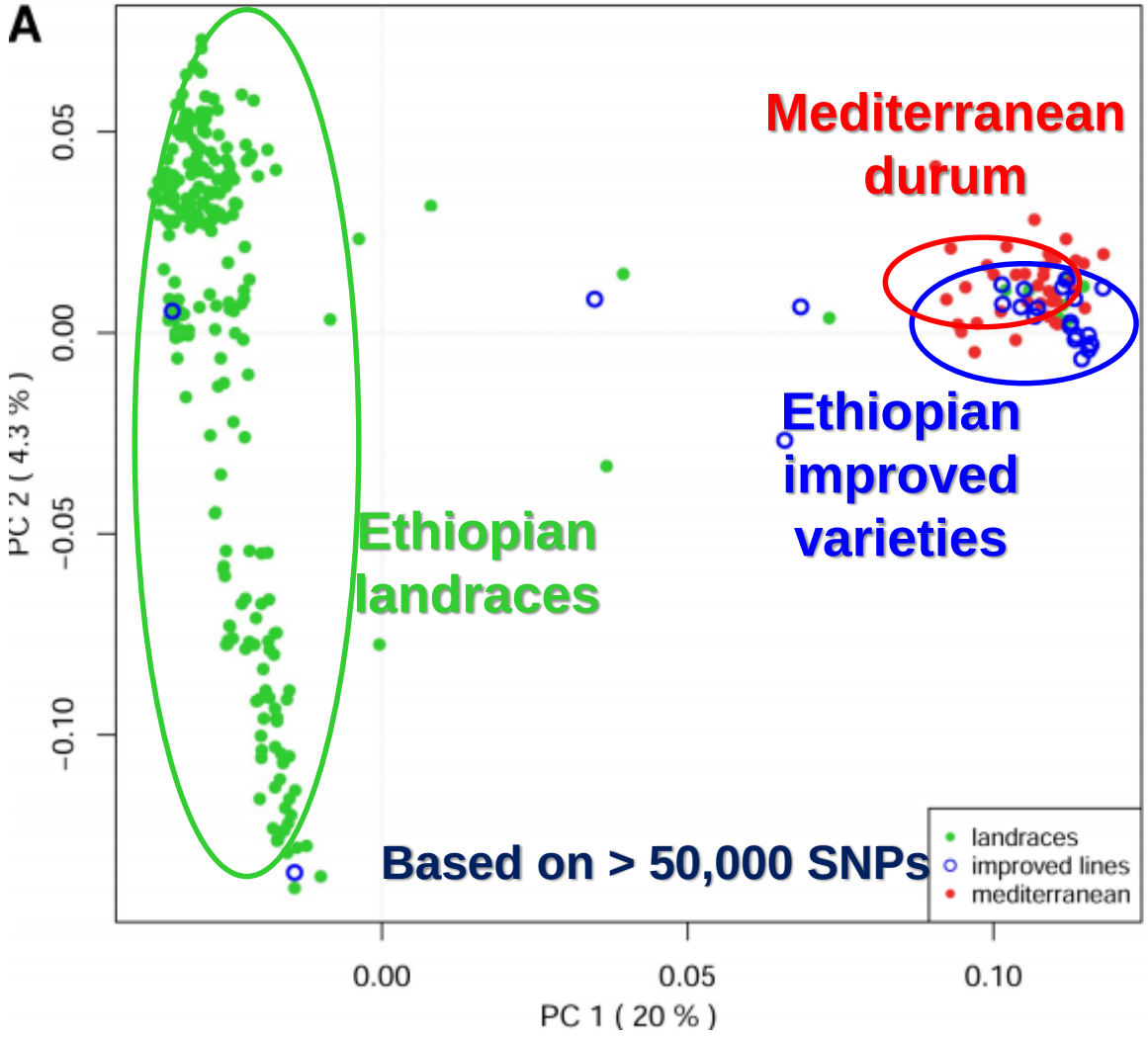
Landraces have broad variation in agronomic traits. Some landraces have interesting stress resistance traits (drought – pests)



90,000 genome-wide SNP markers



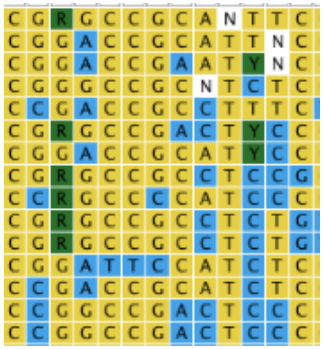
Principal Component Analysis on Molecular Data of the Diversity Panel



MAIN MESSAGES

- Ethiopian landraces are genetically very different
- Most improved varieties are genetically similar
- Ethiopian germplasm is mostly untapped

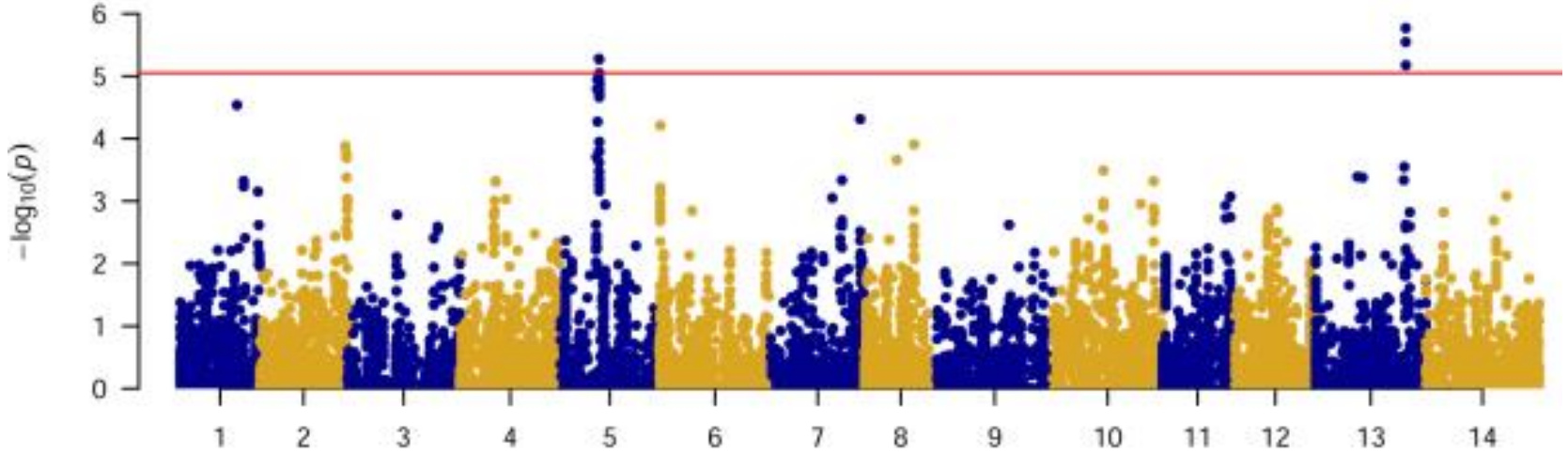
Genome-Wide Association Study (GWAS)



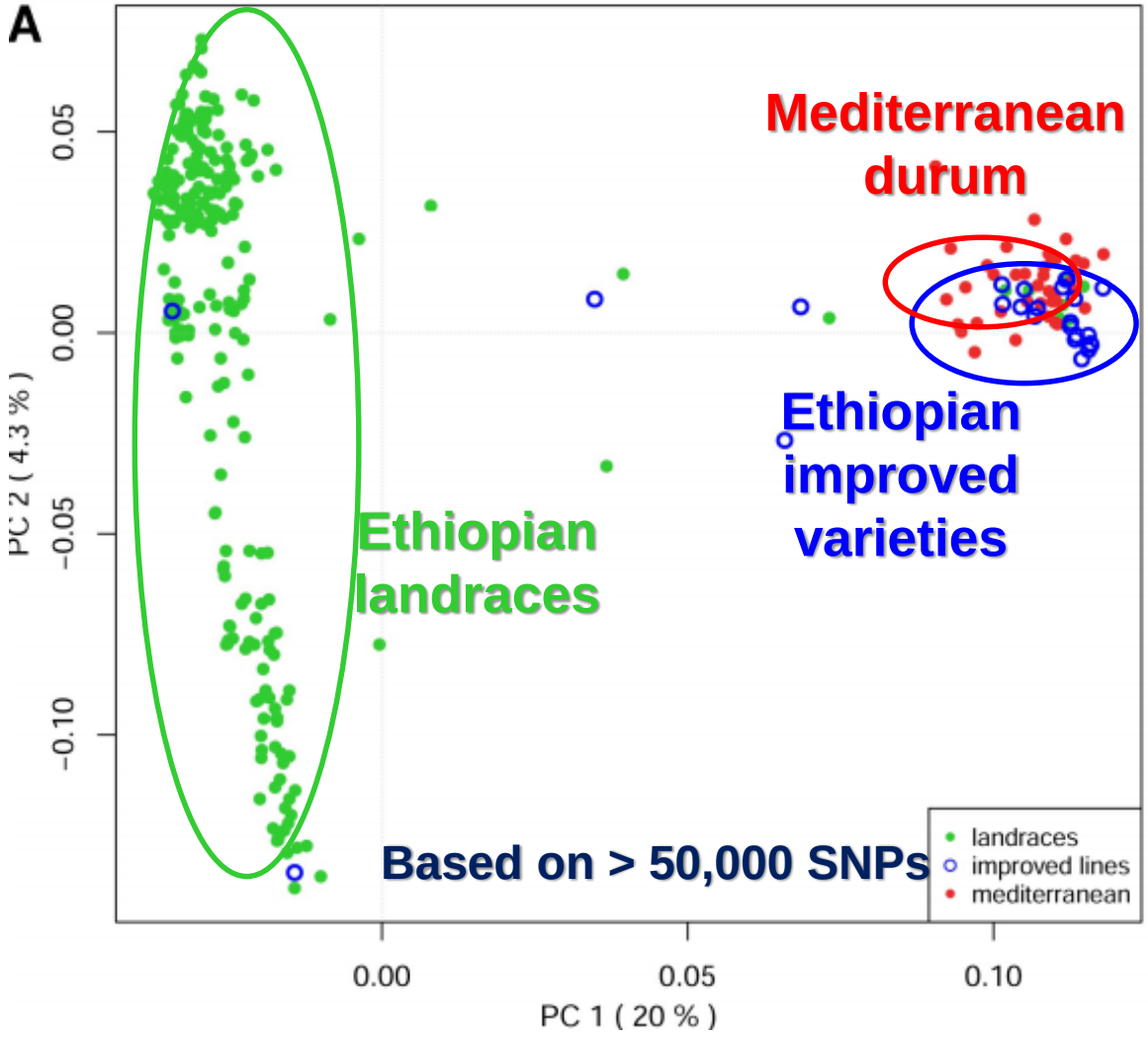
Genome-wide molecular markers

Phenotypic measures / Environmental measures

Statistical analysis - mixed linear models



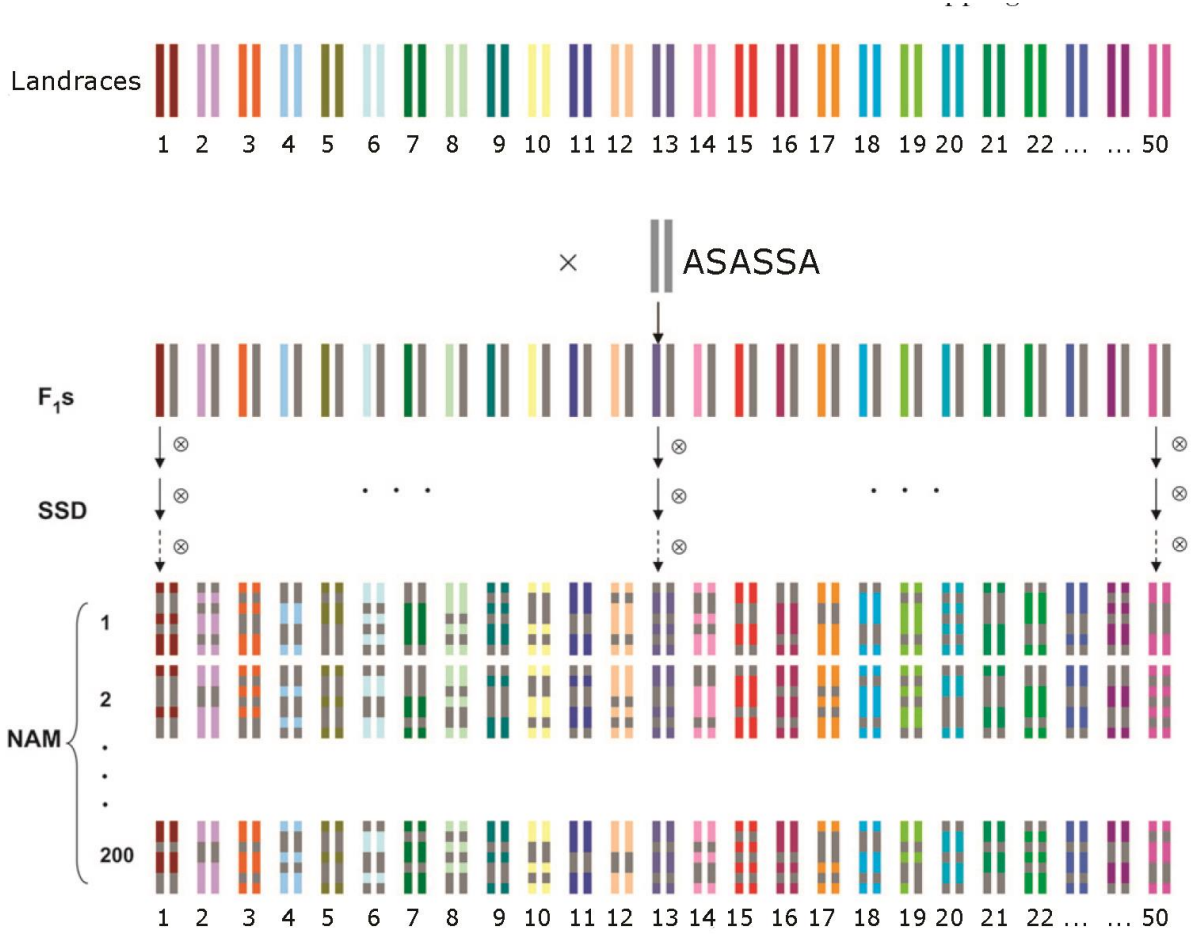
Principal Component Analysis on Molecular Data of the Diversity Panel



MAIN MESSAGES

- Ethiopian landraces are genetically very different
- Most improved varieties are genetically similar
- Ethiopian germplasm is mostly untapped

NAM Population in Ethiopian durum wheat



Modified from Yu et al. 2008

Principal Facts

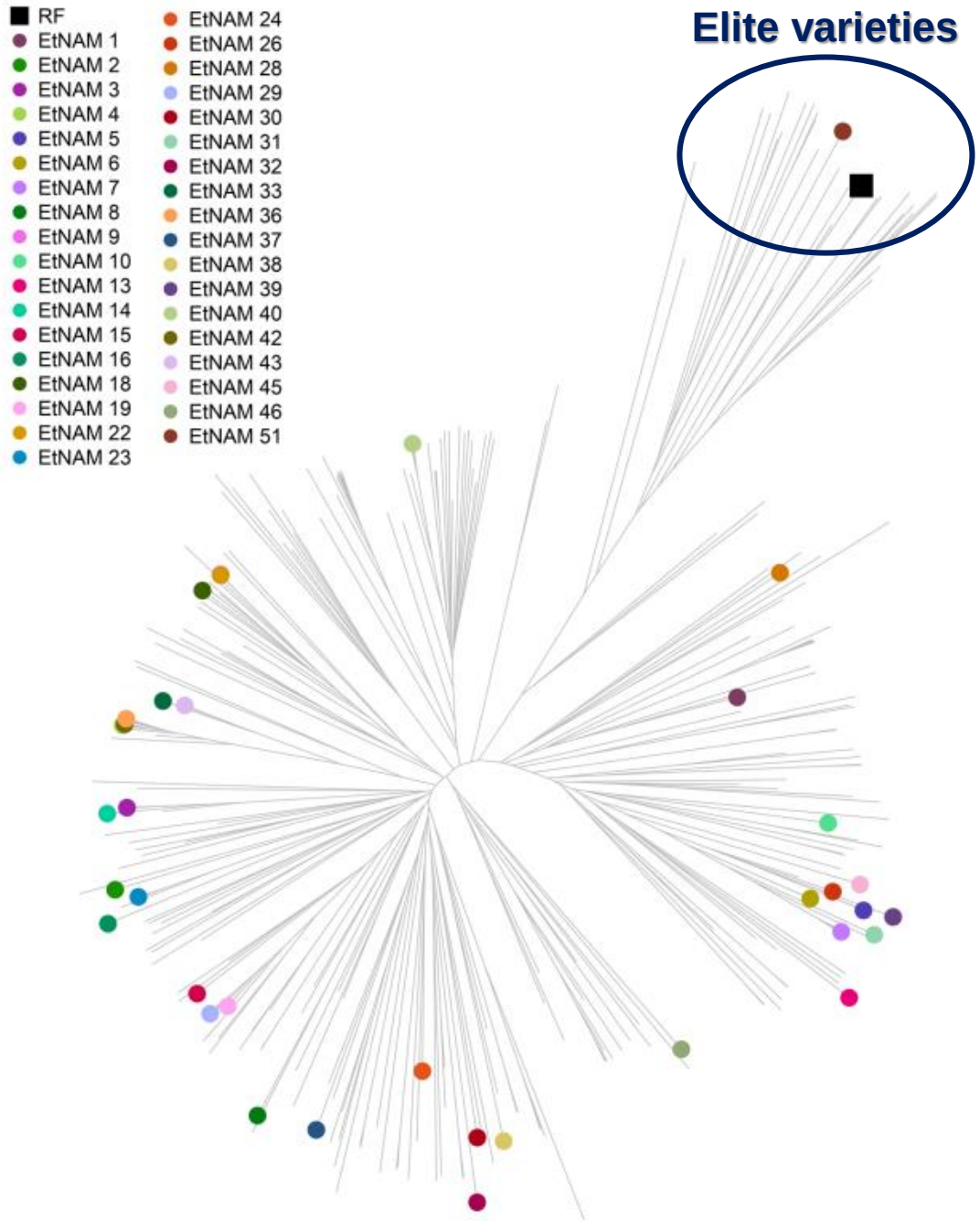
- 50 landraces crossed to an improved variety
- development of 50 RIL families of 80 – 230 lines
- 6,280 F6 lines in available

Founders choice and characterization: phenotypes

Dual Objective:

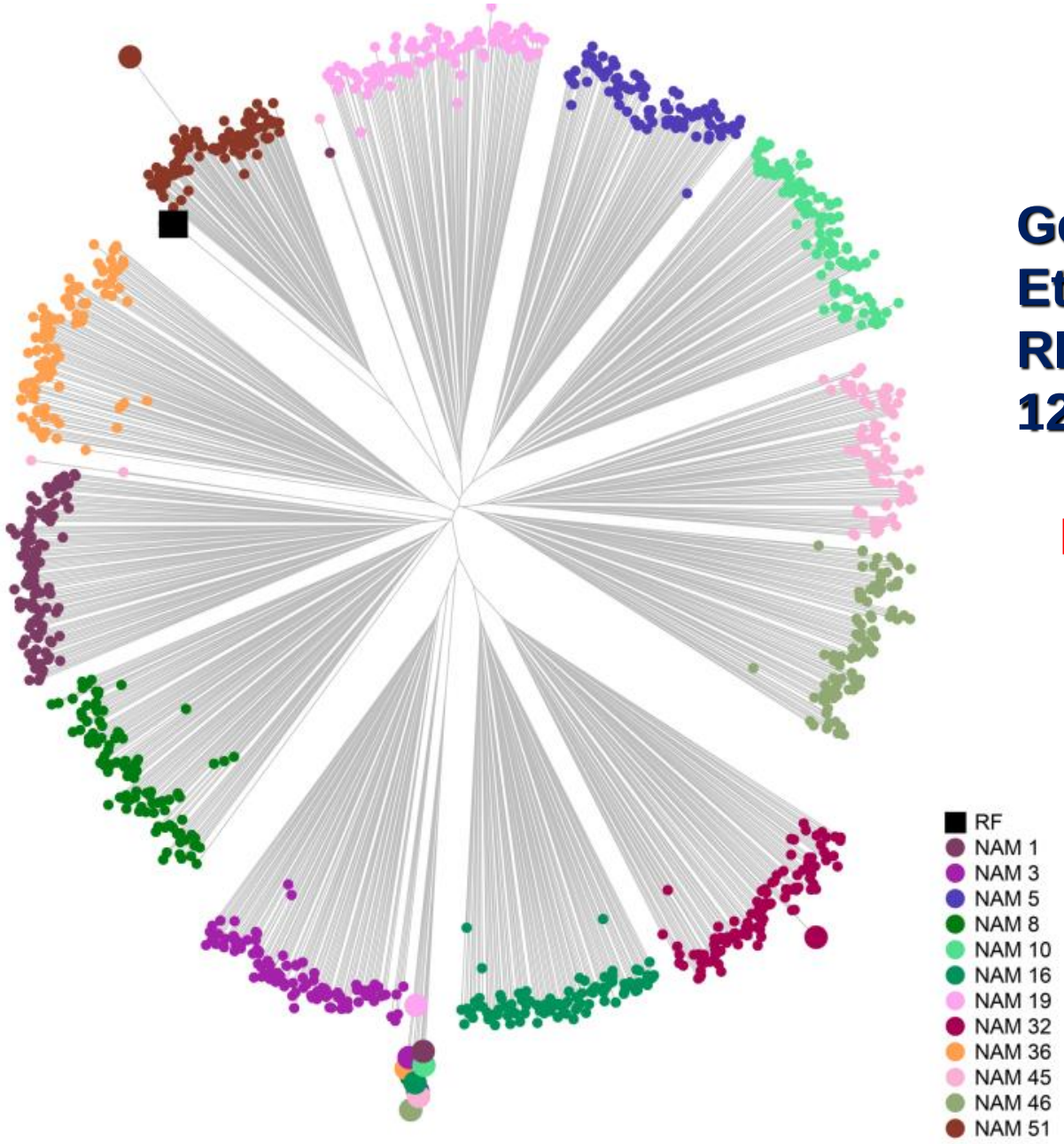
1. **Advanced QTL mapping**
2. **Pre-breeding material**

Some traits deliberately included in the NAM: drought resistance, pest resistance, yield, earliness, rusticity



Founder choice and characterization: genotypes

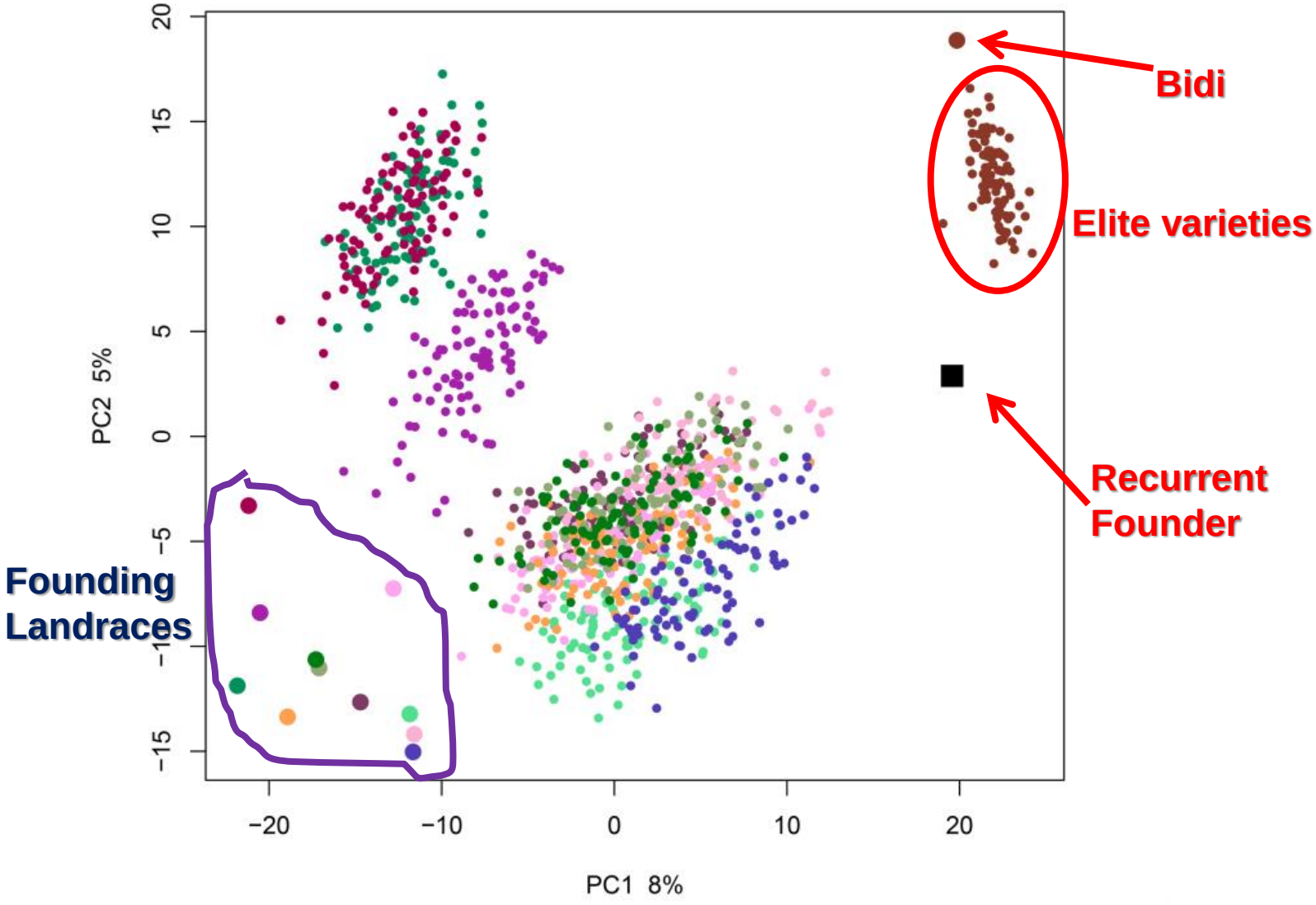
Nj phylogeny representing the molecular diversity of Ethiopian durum wheat



Genotyping of 12 EtNAM Families (100 RILs each) with 12,203 SNPs

Neighbor Joining Tree
12 monophyletic clades

Filling the Gap between Ethiopian Landraces and Elite Germplasm



Phenotypic Characterization and QTL Mapping

2 Locations, 1,300 genotypes, 2 replicates per location two years

10,400 plots

Phenological Traits + Agronomic Traits + Resistance to fungal pathogens + protein content

> 300,000 data points

Septoria tritici blotch
Mycospherella graminicola

Leaf rust
Puccinia triticina

Participatory Evaluation Approches with Local Farmers' Communities

4 Farmers' Traits

Acknowledgments

Yosef Kidane, PhD
Bogale Nigir, MD
Cherinel Aleml, MD

Dejene Mengistu, PhD
Chiara Mancini, PhD

Carlo Fadda, PhD

Sant'Anna
School of Advanced Studies – Pisa

Bioversity International

Mekelle University

ARARI