Setting the Stage: Concepts, Methodologies and Stength of Genetics

A very personal (and biased) view

Data-driven inoovation in breeding programs for local adaptation

Mario Enrico Pè
Scuola Superiore Sant'Anna

SOME IMPORTANT DEFINITIONS - 1

Genetics: the study of biological inheritance

Genotype: the genetic constitution of a CELL or an ORGANISM. It can refer to single or multiple genes

Phenotype: the observable characteristics of a CELL or ORGANISM including the result of tests and/or measurements

Gene: 1. an hereditary unit

2. a functional DNA unit

3. a factor that controls a phenotype and segregates in pedigrees

Alleles: Alternative forms of the same gene

SOME IMPORTANT DEFINITIONS – 2

Mutation: a heritable gene or chromosomal or genomic alteration

Mutant: 1. a biological entity with a change due to a mutation

2. a "changed" gene

Segregation: the distribution of different alleles in daughter cells at meiosis

Recombinant: a gamete that contains a combination of allelles that is different from the combination of the parents

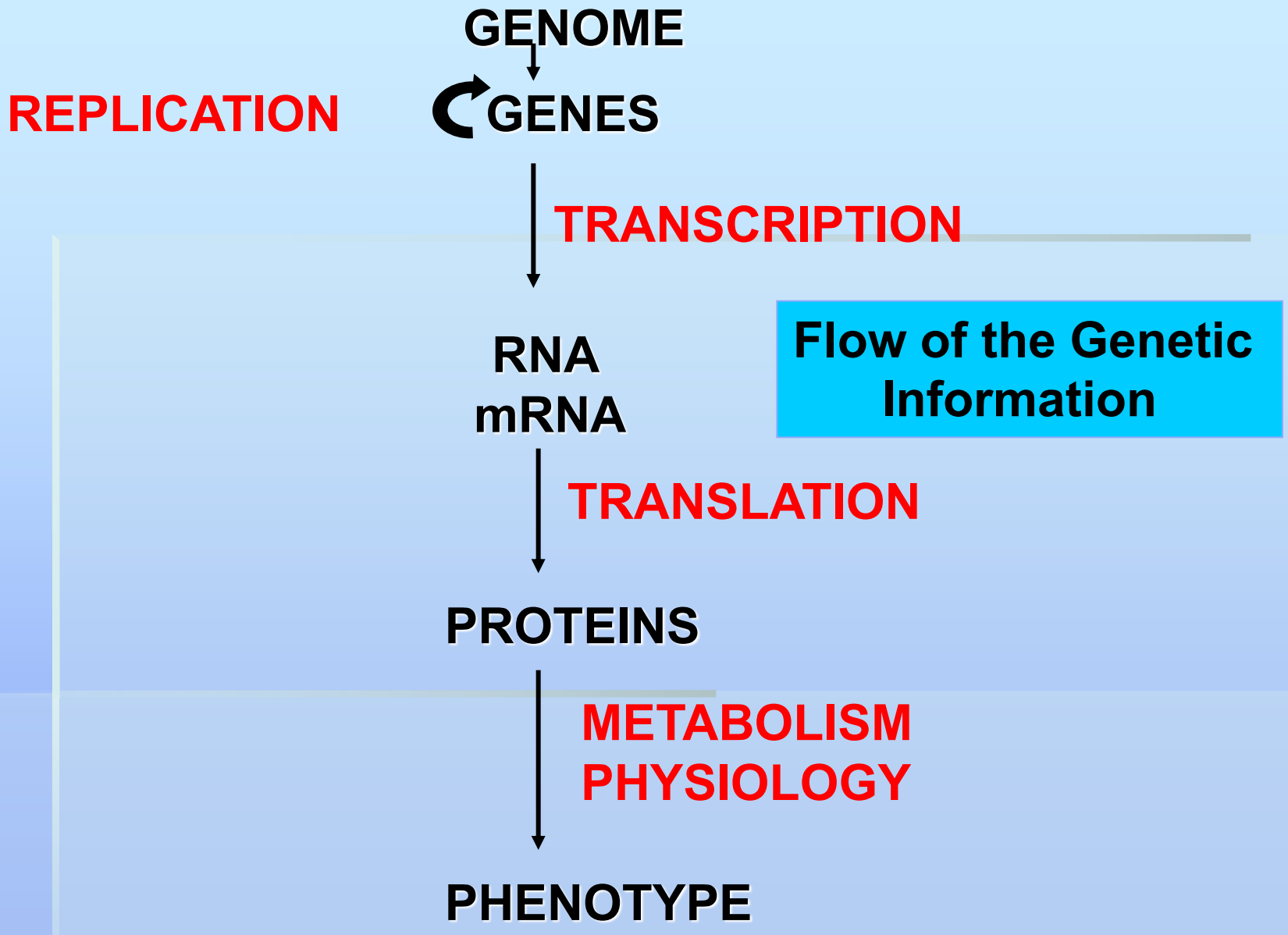
GENOTYPE- PHENOTYPE RELATIONSHIPS: STILL THE QUESTION

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

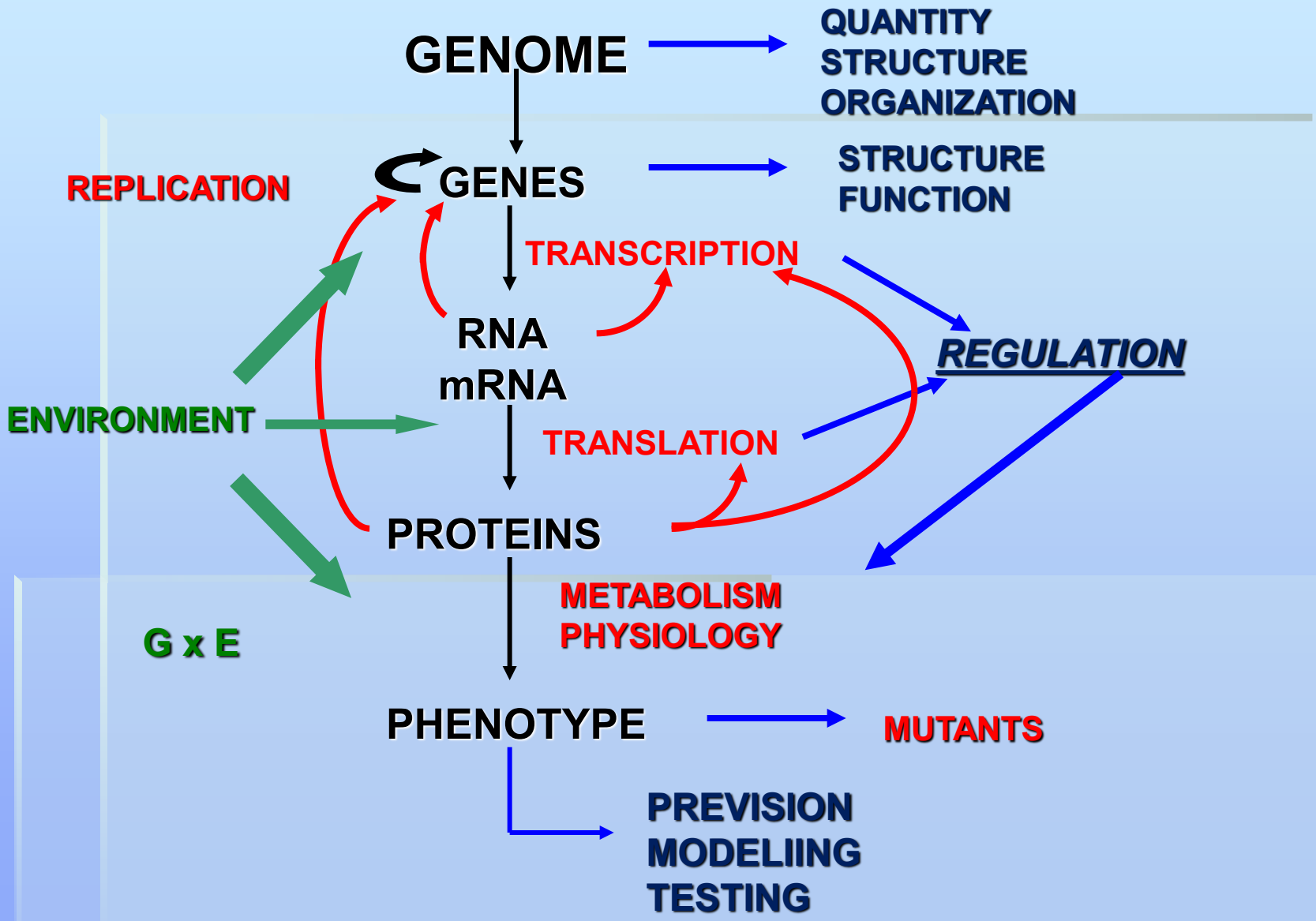
DIFFERENT SCHOOLS

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

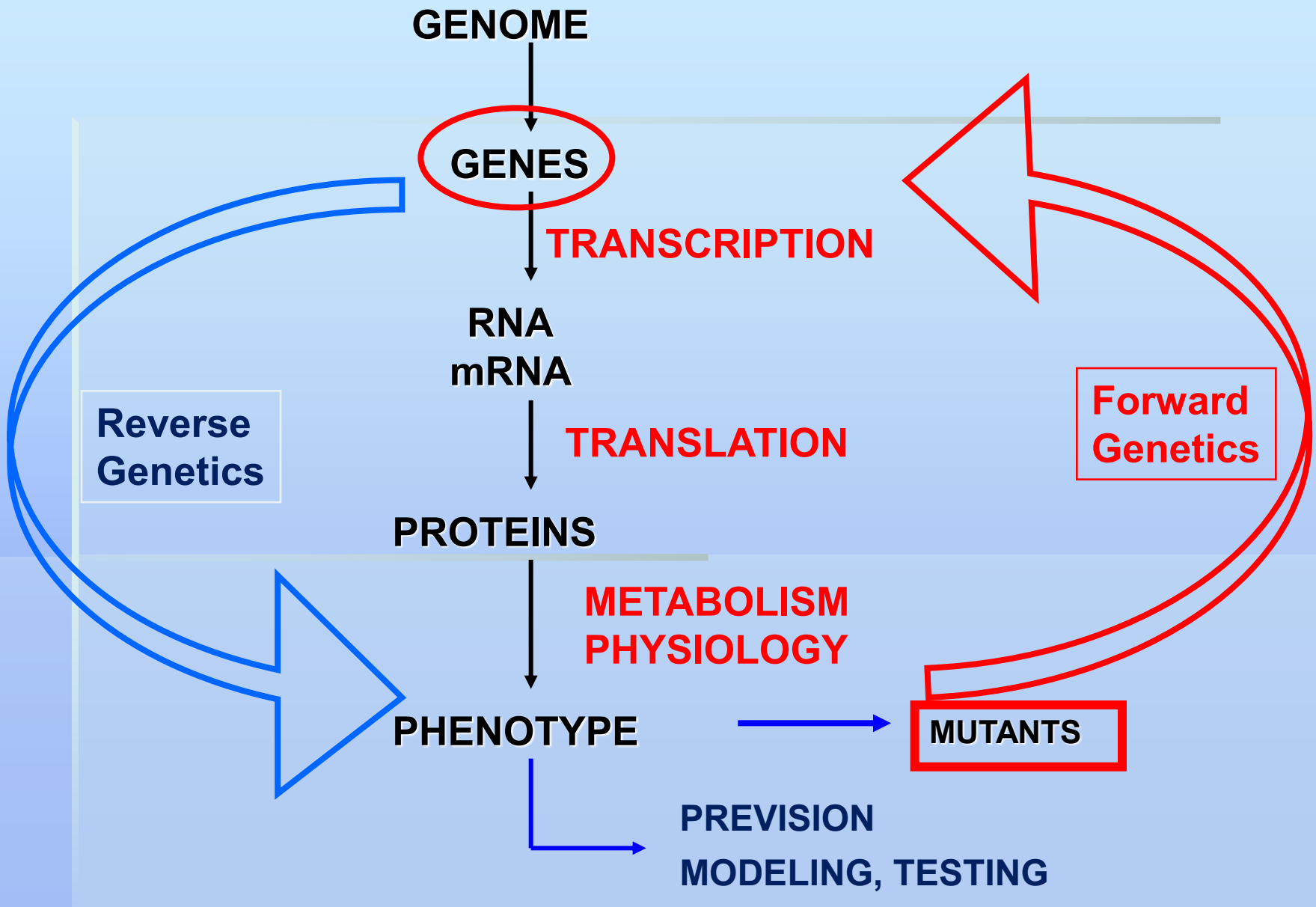
2. MENDELIAN SCHOOL
Approach: Reductionist
Methods: Mendelian Genetics, MolecularGenetics


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

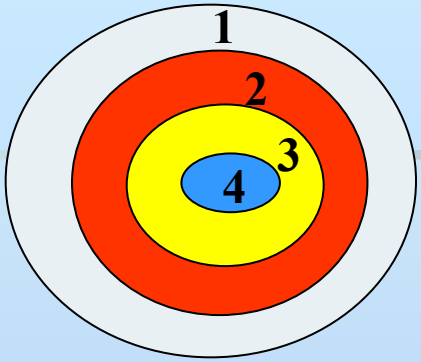
The Flow of the Genetic Information



Genetic Approaches



EXAMPLE of FORWARD GENETICS: FLOWER DEVELOPMENT in *Arabidopsis thaliana*



Arabidopsis thaliana

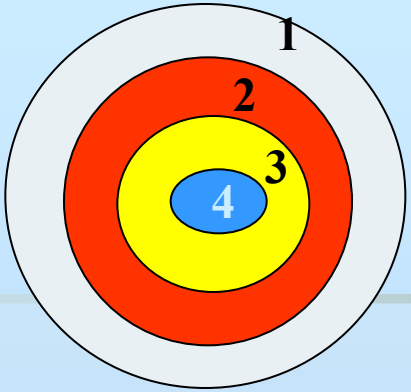
whorl = a circular arrangement of leaves or flowers or other organs radiating from a single node

Concentric whorls

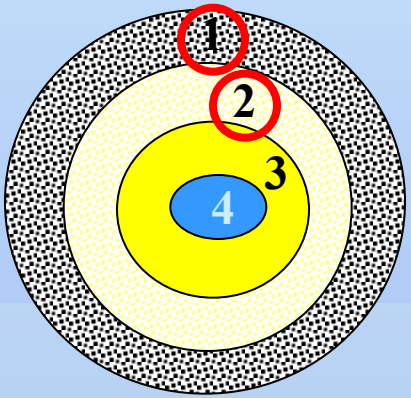
1. **Sepals**
2. **Petals**
3. **Stamens**
4. **Carpels**

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

Normal flower

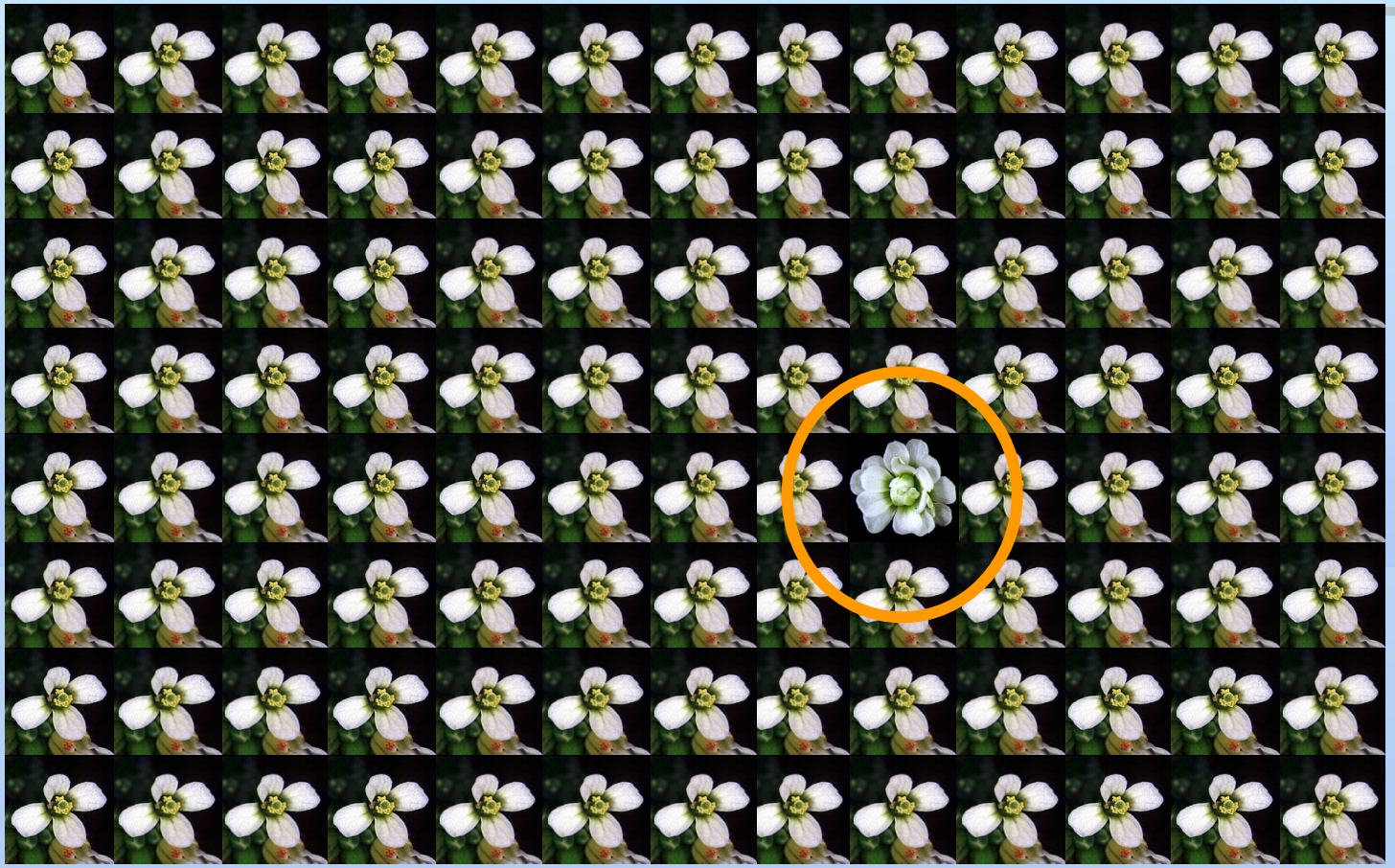


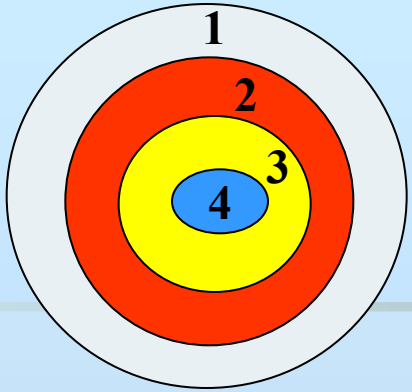
Mutant *apetala 1*



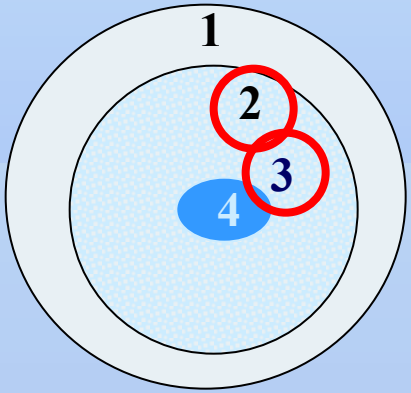
carpels stamens stamens carpels

Program of mutagenesis and mutant selection



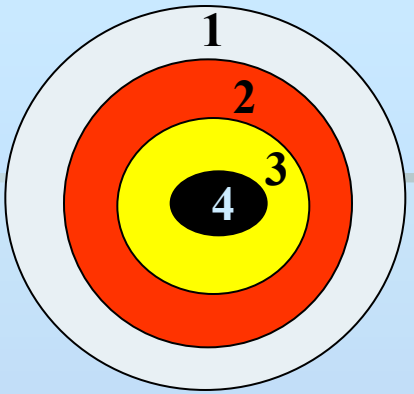


Normal flower

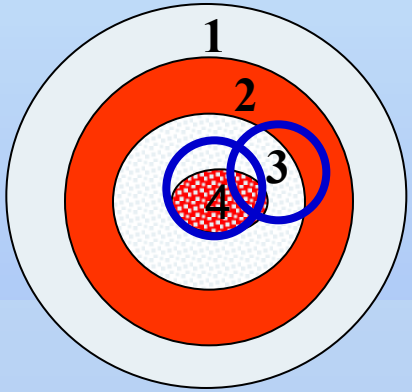


Mutant *pistillata* Sepals sepals carpels carpels

Normal flower



Mutant *agamous*



sepals petals petals petals

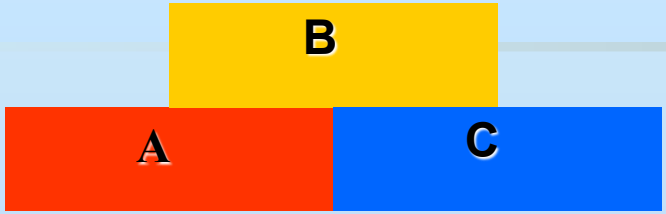
1. Recessive mutations
2. Mutations in a single gene
3. Mutations affecting two adjacent whorls
4. Homeotic mutations

(Homeosis is the transformation of one body part into another, arising from mutation in or mis-expression of specific developmentally critical genes)

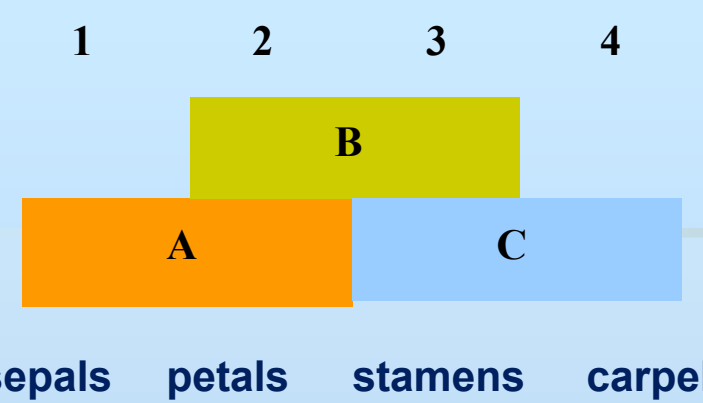
normal organs which develop in wrong places

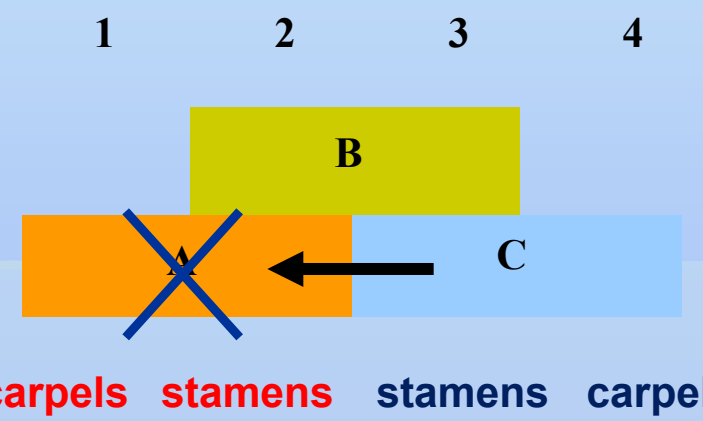
ABC Model

1 2 3 4

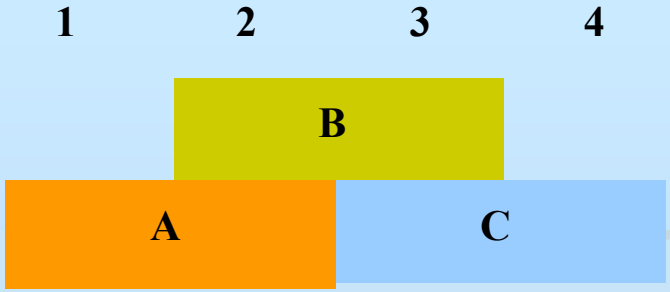


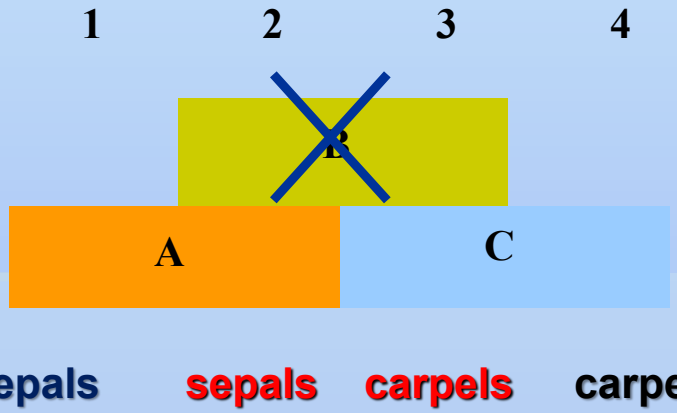
sepals petals stamens carpels



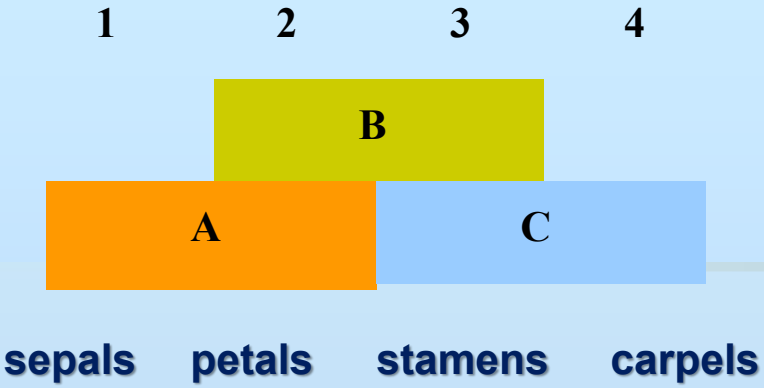


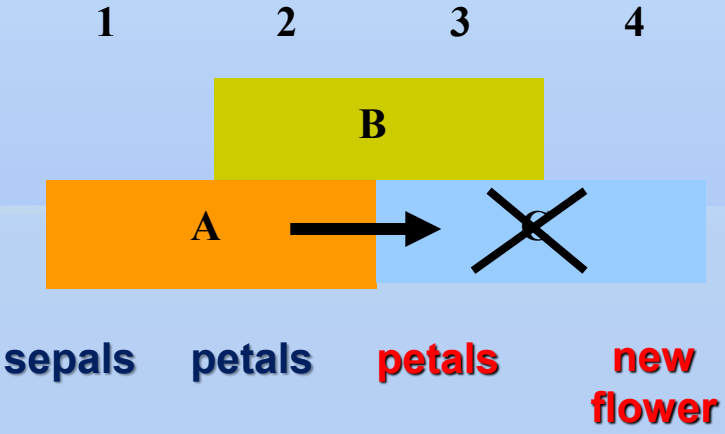
Conclusion: *Apetala 1* is a class A gene





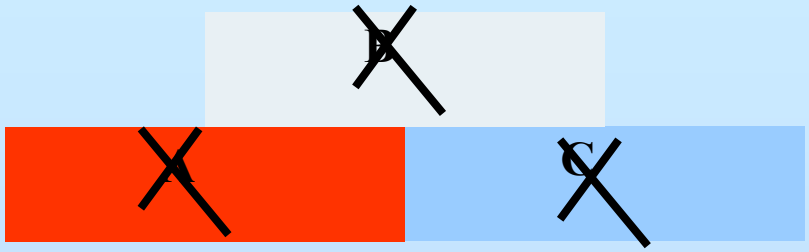
Conclusion: *Pistillata* is a class B gene





Conclusion: *Agamous* is a class C gene

1 2 3 4



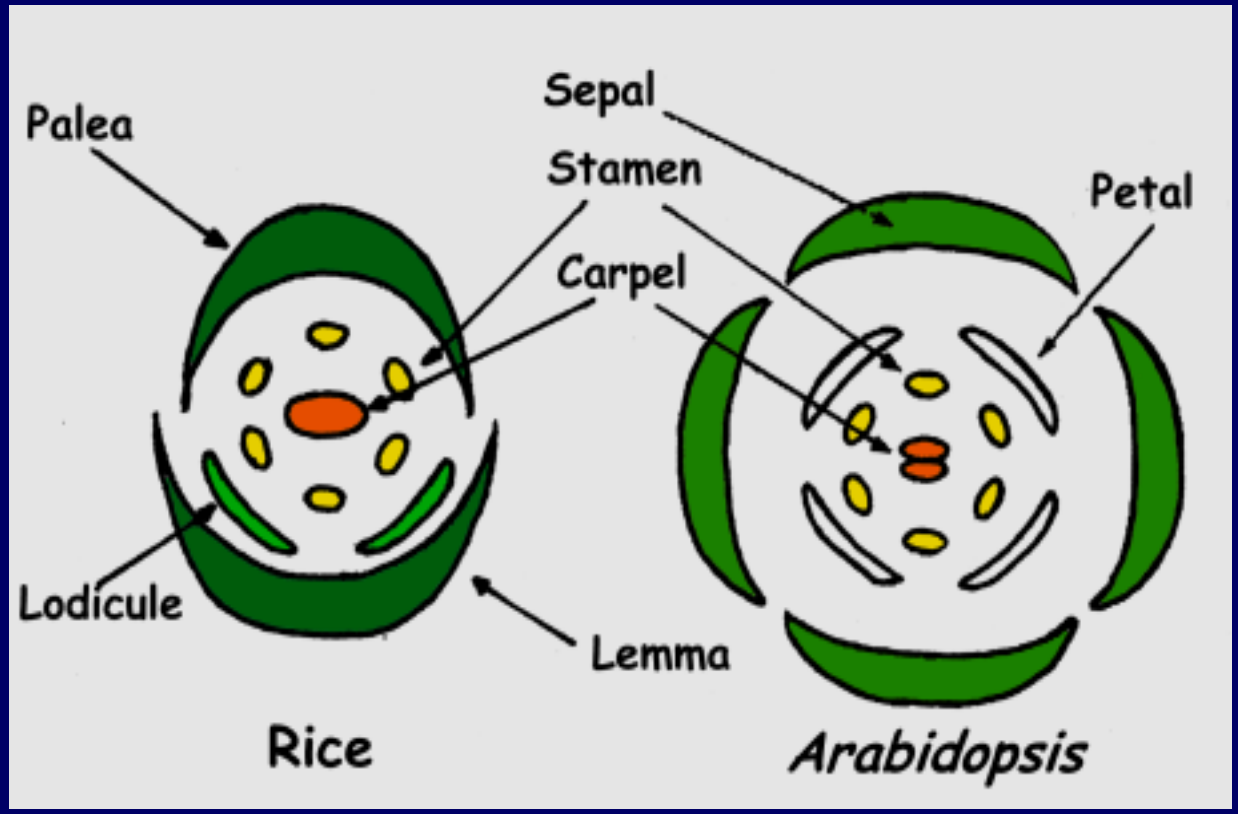
leaf leaf leaf leaf

The triple mutant abolishes formation of all floral organs

Conclusion: the formation of a leaf is the "basal program"
The flower is a transformed leaf

Arabidopsis

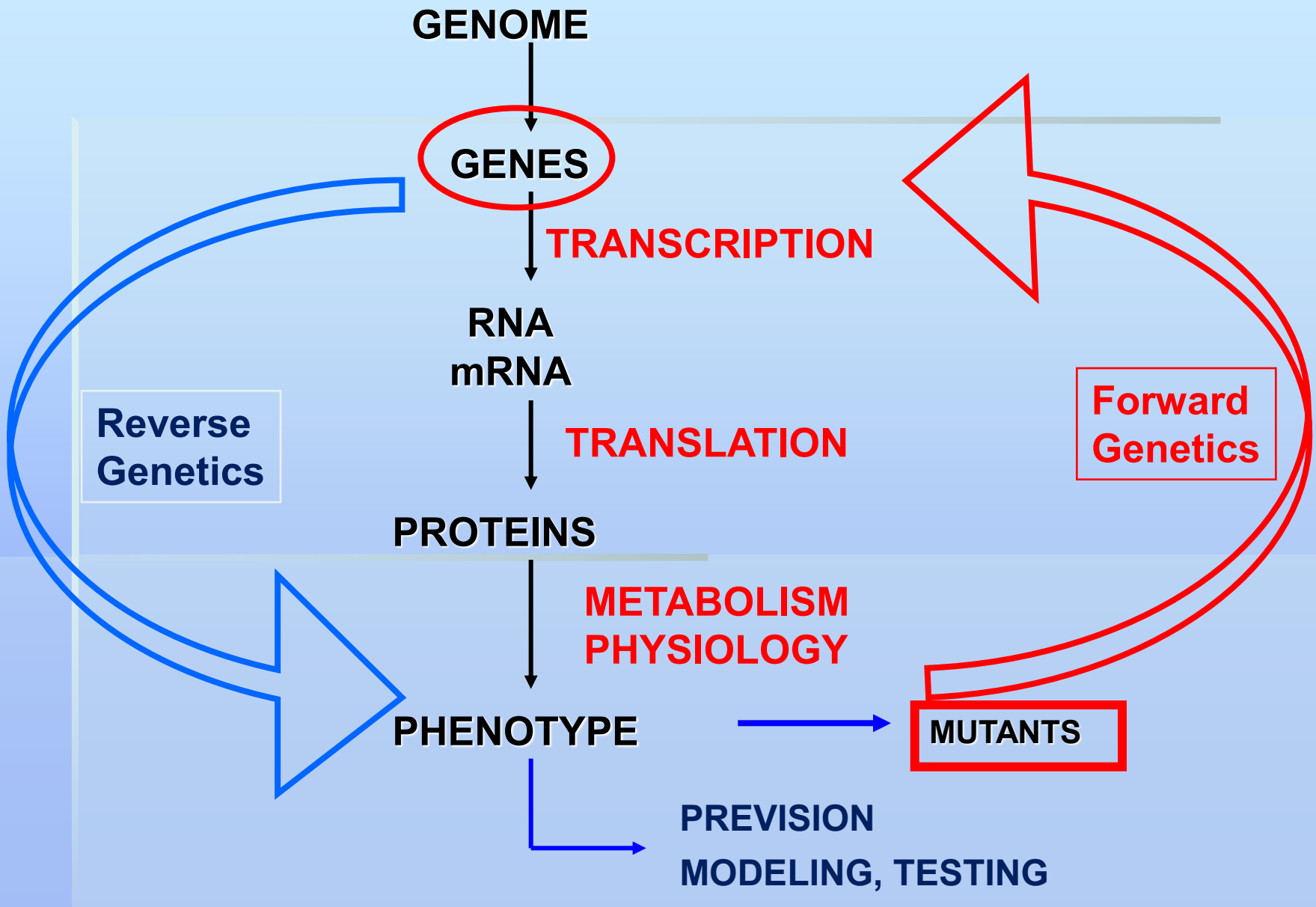
Rice


Palea
Sepal
Stamen
Petal
Carpel
Lodicule
Lemma
Rice
Arabidopsis

Link Between Forward Genetics and Reverse Genetics

- Homeotic genes, coding for transcription factors, control 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 whether they are involved in flower development as well

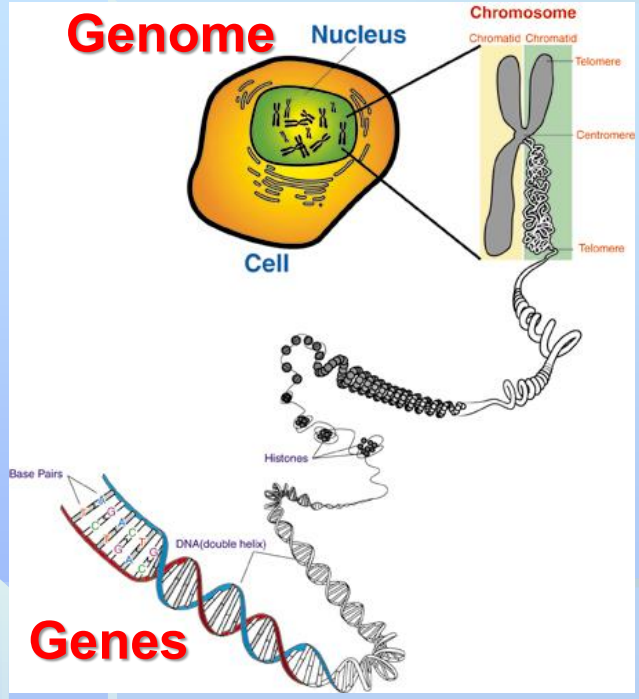
Genetic Approaches



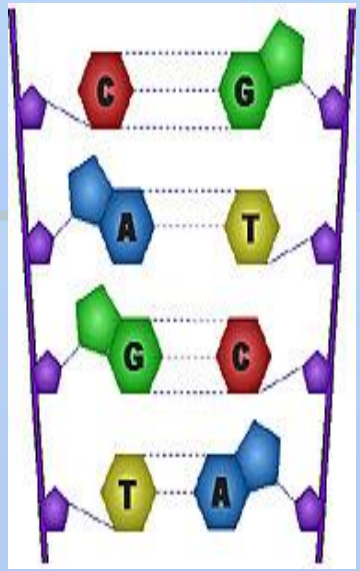
GENOME

Whole of the GENETIC MATERIAL - DNA -

in a cell, in an organism, or in a species



The functional potential of an organism



Genome Complexity
Humans = 3,520 Mb
Durum wheat = 17.000 Mb

Genotype **Traits (Phenotype)**

GENETIC MANTRA: NO genetic variation NO Genetics

- Genetics addresses differences among individuals of a species (Sturtevant & Beadle 1939)
- Variability among individuals is crucial for a species to adapt to the environment and evolve, either by selection and / or by random drift
- Selection (natural and artificial) works on PHENOTYPIC differences

MUTATIONS

1. Source of new genetic variability
2. Cornerstone of genetic analysis
3. Studying aberrant phenotypes may lead to the discovery of wild type function of a gene

Mutation and Genetic Variability

Mutations produce genetic variability by changing:

1. **Structure of the genetic material, i.e. chromosomal rearrangements**
2. **Amount of the genetic material, *i.e.* polyploidy, gene duplication, gene loss**
3. **Modification of DNA sequences, *i.e.* point mutations**
4. **Mutations are ubiquitous in the genome**
5. **Mutations occur in every generation**

MUTATIONS

Mutations can be divided into three main types

1. Chromosome mutations

   Changes in chromosome structure

2. Genome mutations

   Changes in chromosome number

3. Single-gene mutations

   Relatively small changes in DNA structure that occur within a particular gene

Mutations and Phenotypic Effects

- Only a small portion of mutations leads to phenotypic changes the majority are SILENT but still provide genetic variation
- Only those which cause phenotypic effects provide phenotypic variability on which natural selection can act
- Most mutations produce negative phenotypic effects
- Linking mutation, *i.e.* genetic variation, to phenotypic effects IS a major problem

Mutations and Phenotypic Effects

- A minor portion of mutations leads to phenotypic changes
- Only those which cause phenotypic effects provide phenotypic variability on which natural selection can act
- Most mutations produce negative phenotypic effects
- Determining phenotypic effects IS still a major problem

Classification of mutations based on molecular changes

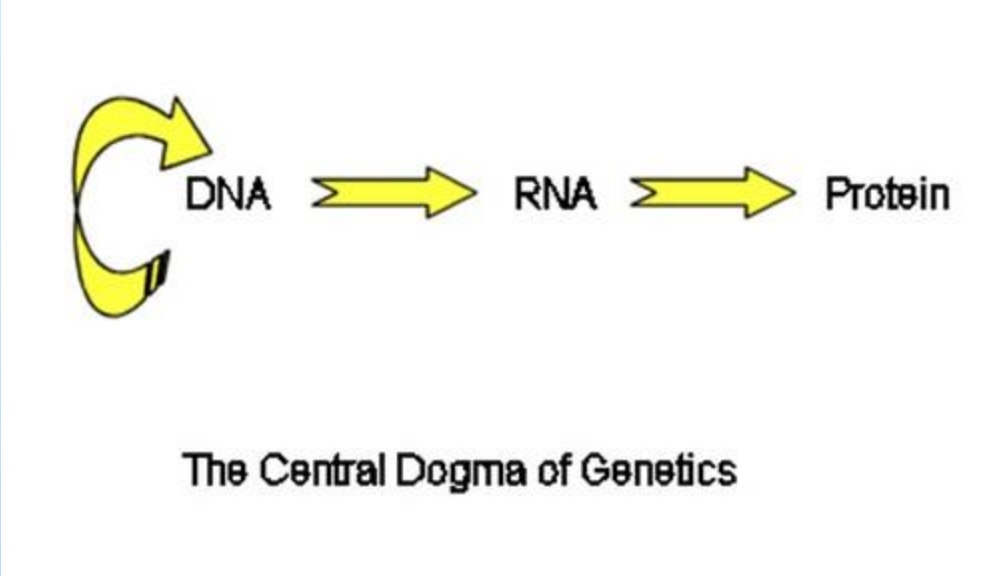
Genetic Mutation

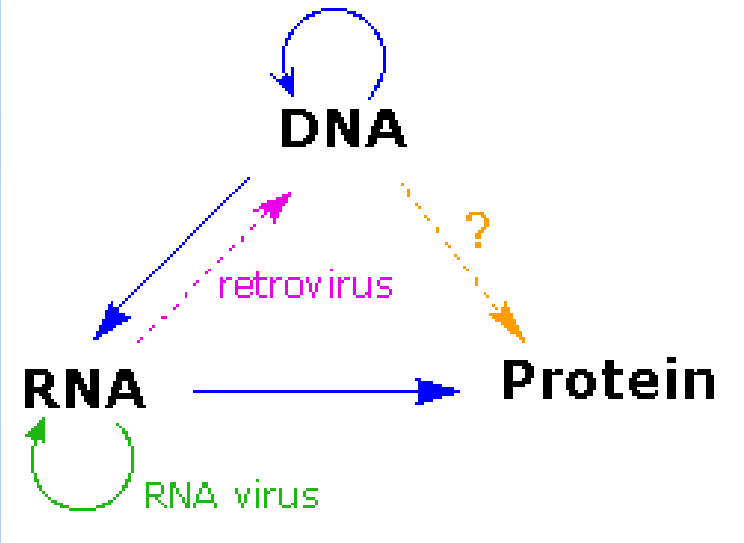
- Base Subtitution
- Insertion / Deletion of nucleotides

They are caused by many factors including environmental causes (radiation and mutagenic chemicals) and their position is almost random

MOLECULAR BIOLOGY BOX

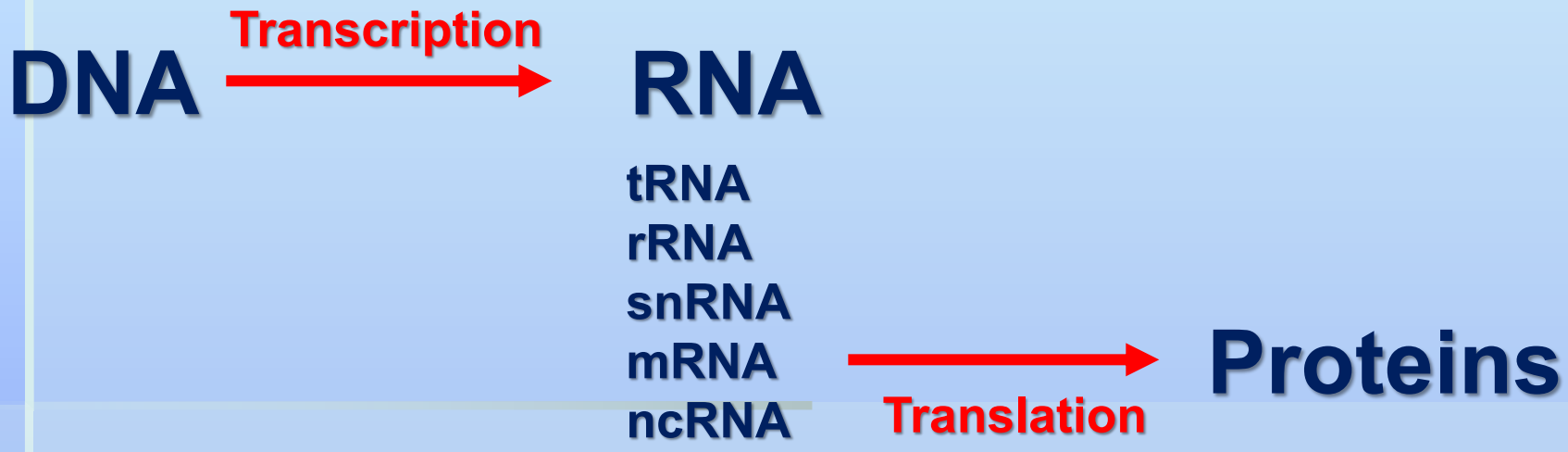
CENTRAL DOGMA of BIOLOGY





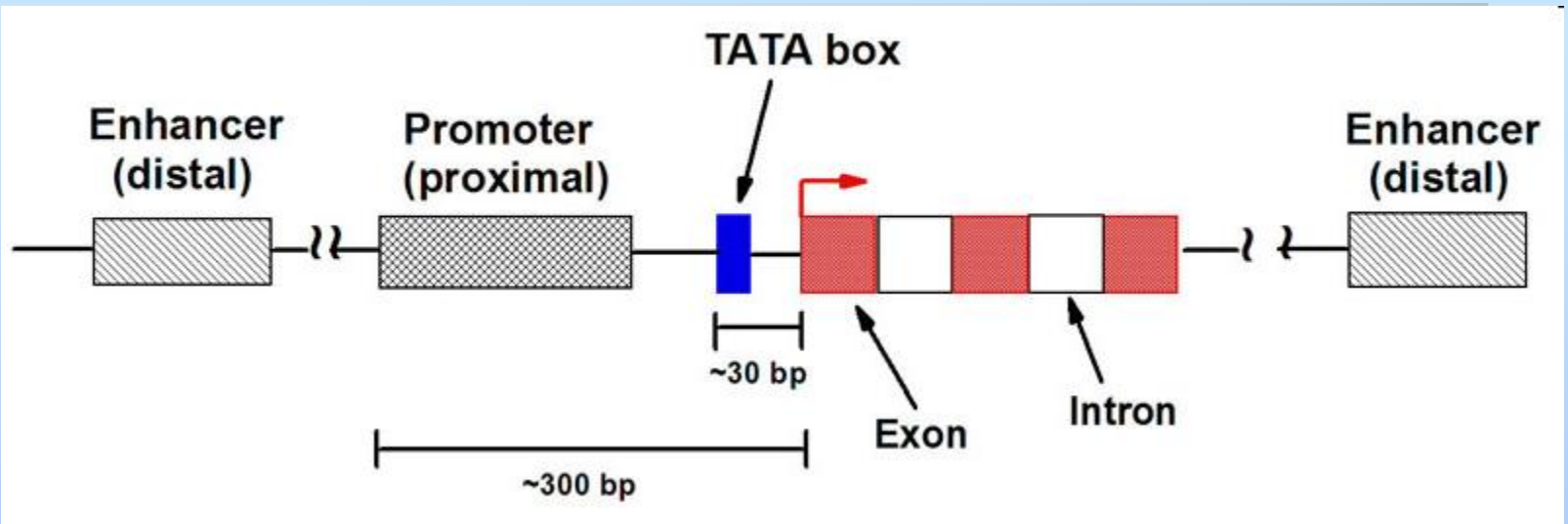
Transcription is the primary gene function

The synthesis of RNA molecules using DNA strands as the templates so that the genetic information can be transferred from DNA to RNA

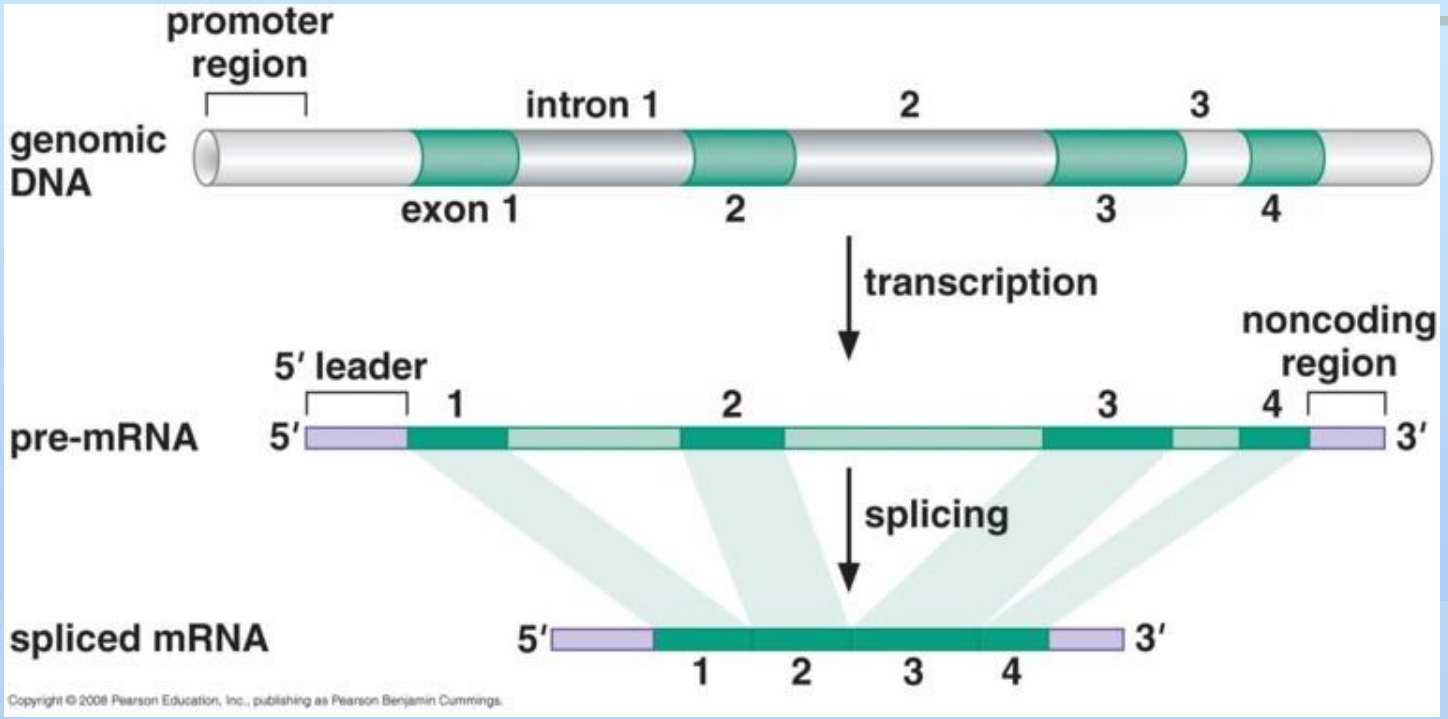


EACH STEP REPRESENTS AN AMPLIFICATION OF THE GENETIC INFORMATION

Eukaryotic Gene Structure



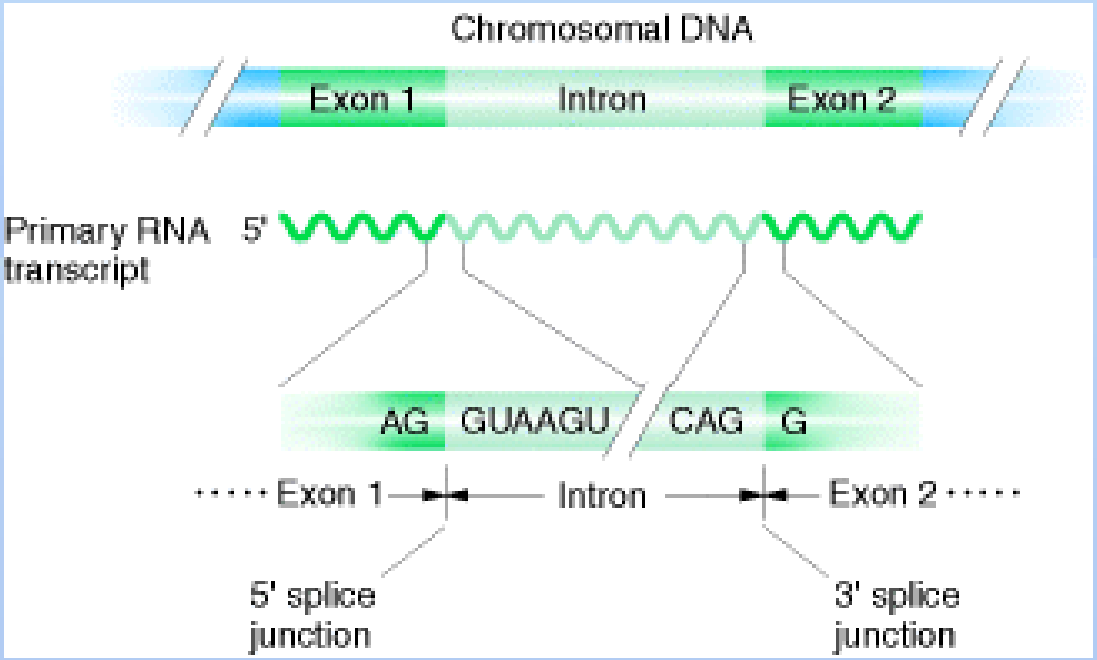
Eukaryotic genes contain introns which are spliced to form mature mRNA



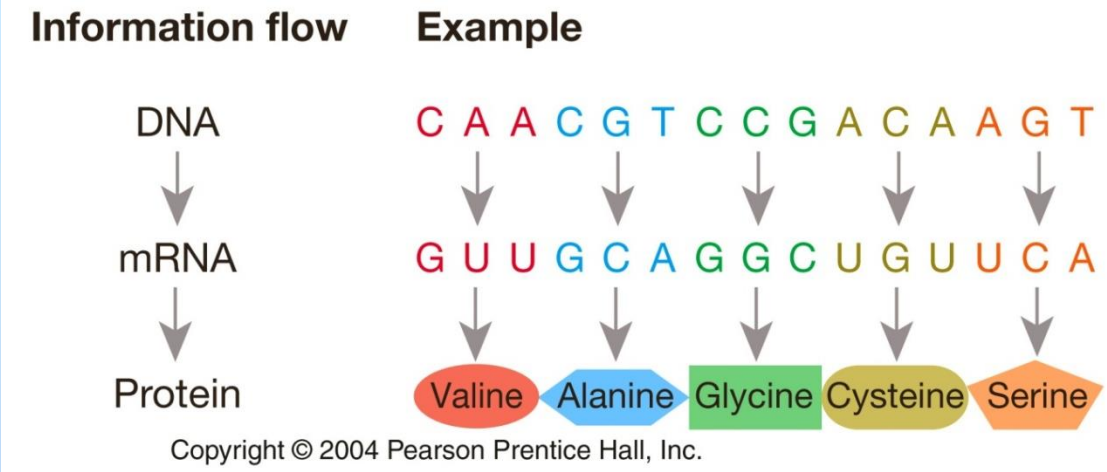
Splice Sites

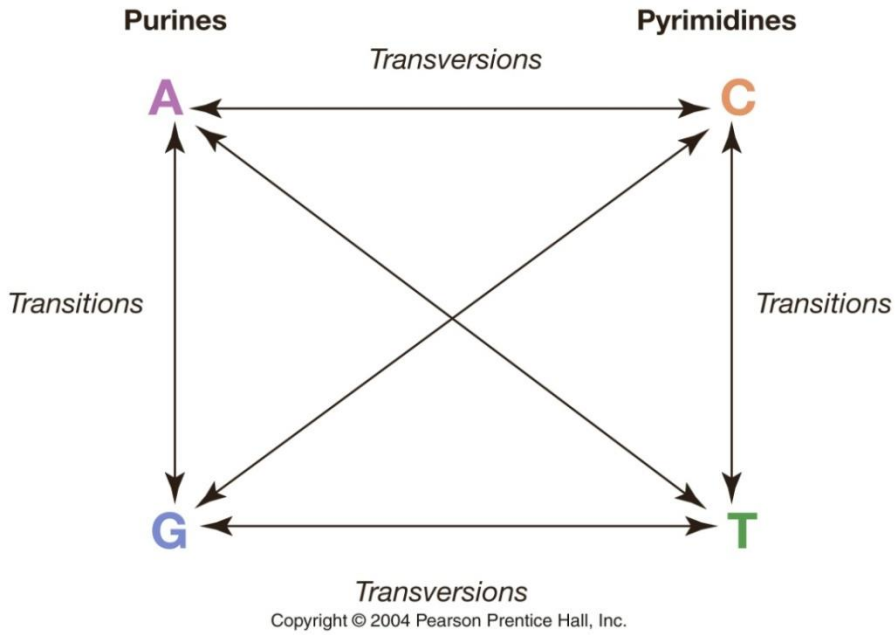
- Conserved splice sites are shared by both the exon and the intron

- Different signals on the donor site (3') and on the acceptor site (5')



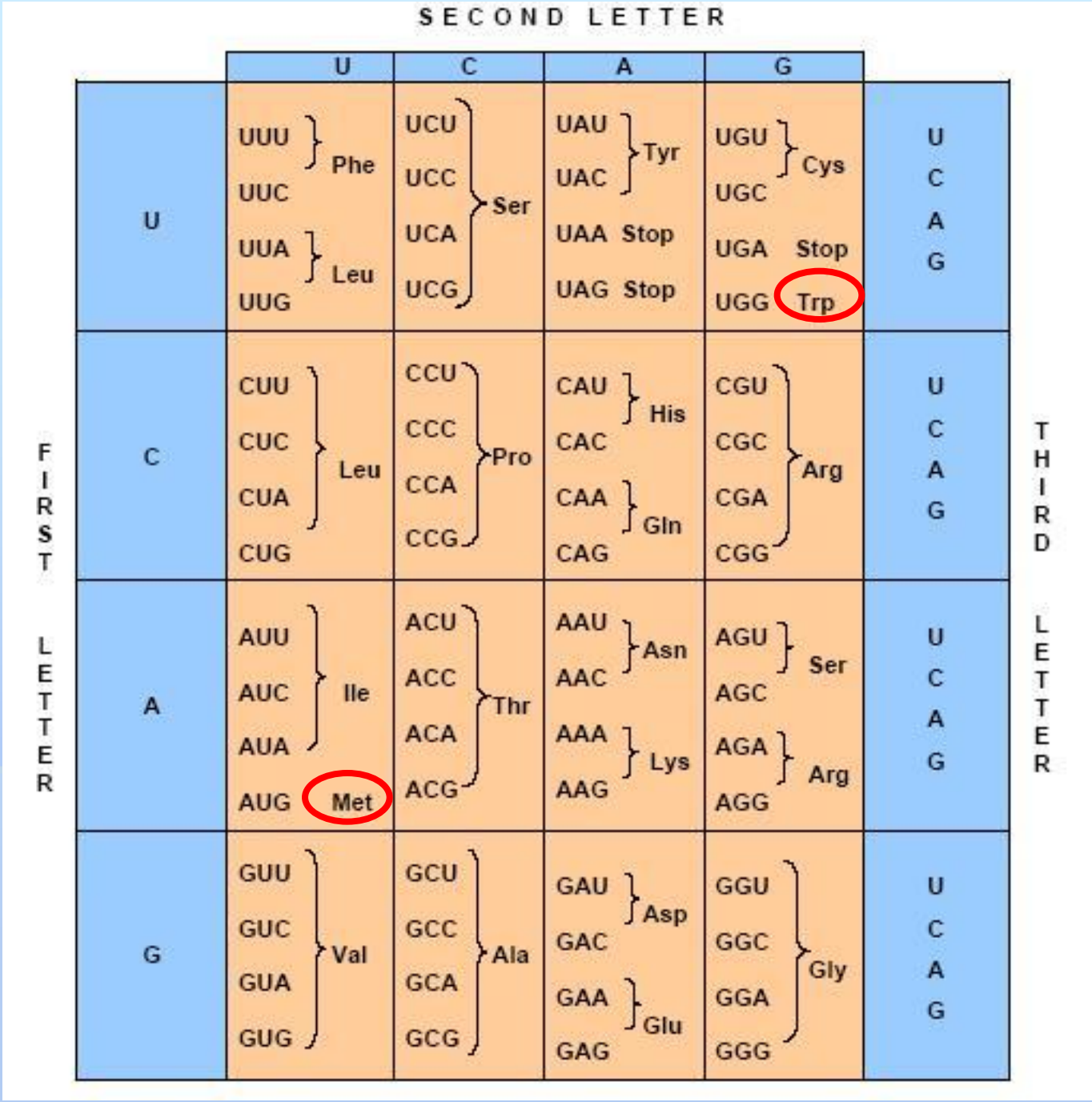
Production of New Alleles in CDS by nucleotide substitutions





- **Transitions are more frequent than transvertions**
  - **likely because transitions cause a smaller modification to the DNA molecule that can go un-noticed by the DNA repair machinery**
  - **mutations that DO NOT determine an a.a. change are called “silent”**

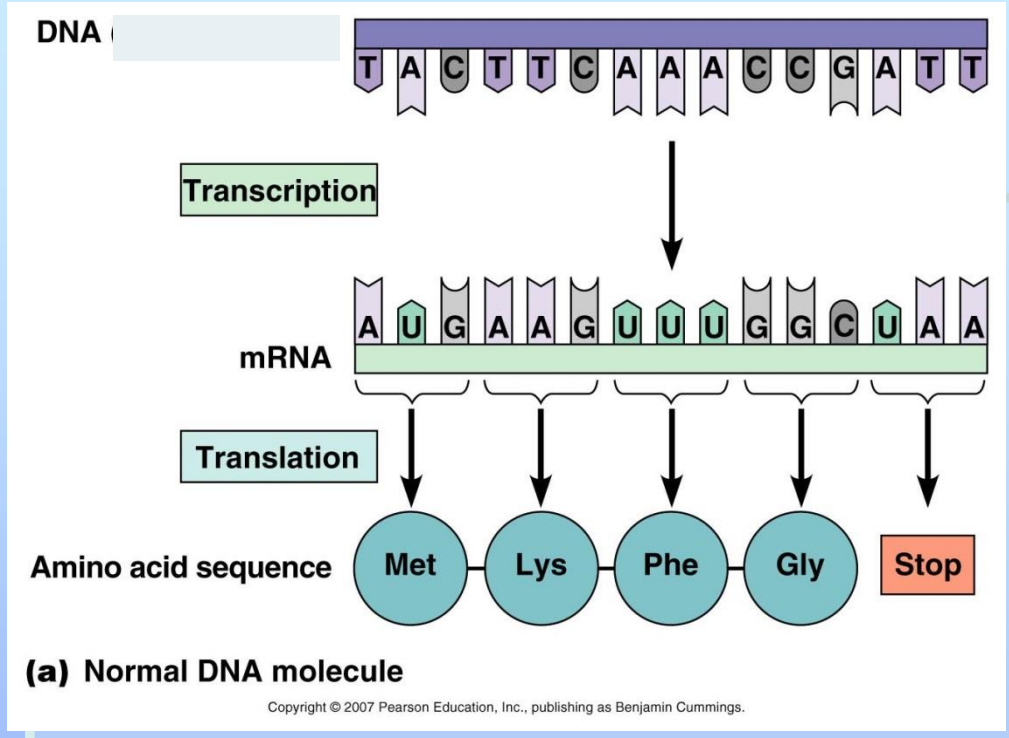
Genetic Code



Degeneracy of the code affects phenotypic effects of mutations

- **Degeneracy of the genetic code implies that EVEN in CDS a large number of mutation DOES NOT determine an a.a. change**
- **Those mutations are called *silent mutations***
- **Silent mutations arise from base sustitution at the third base of the codon. Still they produce genetic variation and therefore…….**

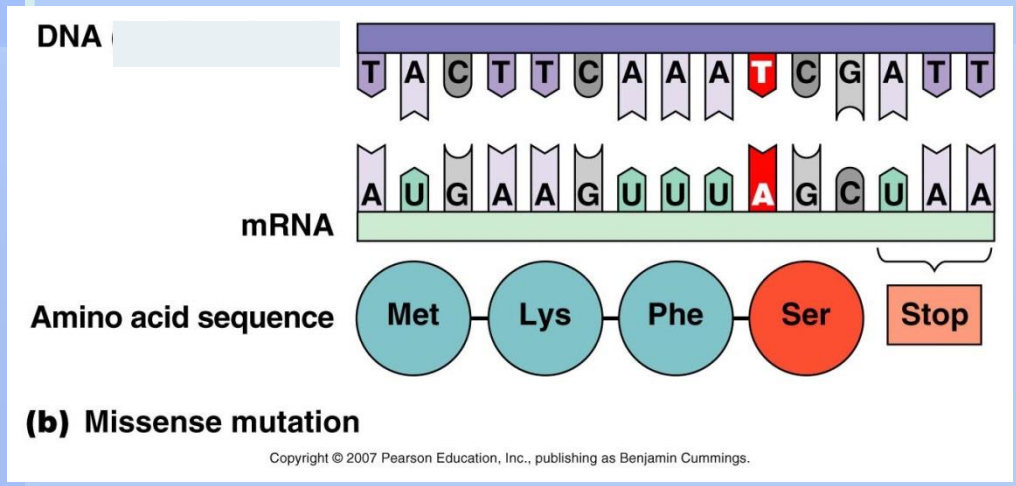
Single Nucleotide Substitution in Coding Sequence



Template strand (complementary to RNA)

RNA

Protein

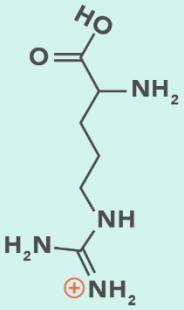


Transition G – A in first base
Different a.a.
Missense mutation

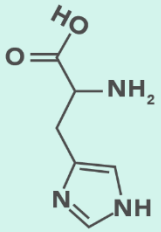
A. Amino Acids with Electrically Charged Side Chains

Positive

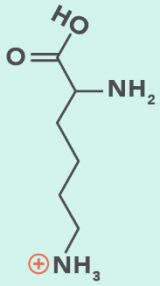
Arginine (Arg) R



Histidine (His) H

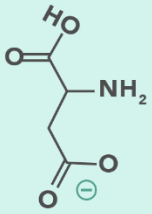


Lysine (Lys) K

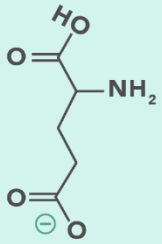


Negative

Aspartic Acid (Asp) D

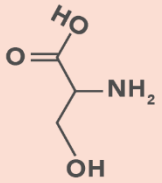


Glutamic Acid (Glu) E

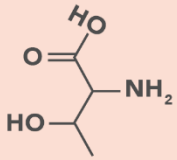


B. Amino Acids with Polar Uncharged Side Chains

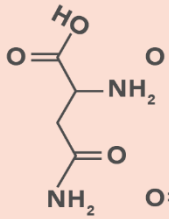
Serine (Ser) S



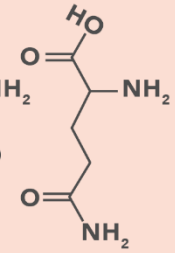
Threonine (Thr) T



Asparagine (Asn) N

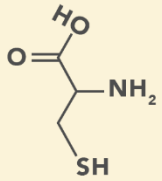


Glutamine (Gln) Q

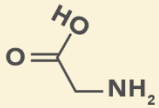


C. Special Cases

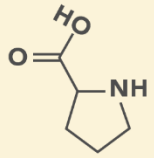
Cysteine (Cys) C



Glycine (Gly) G

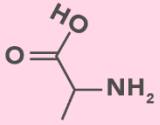


Proline (Pro) P

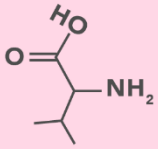


D. Amino Acids with Hydrophobic Side Chains

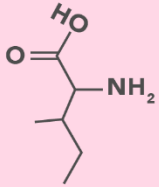
Alanine (Ala) A



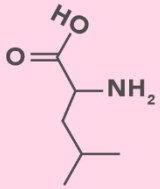
Valine (Val) V



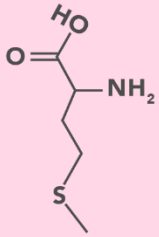
Isoleucine (Ile) I



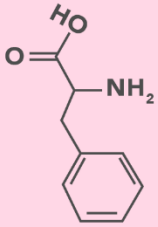
Leucine (Leu) L



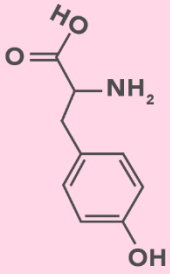
Methionine (Met) M



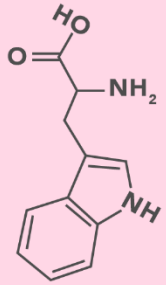
Phenylalanine (Phe) F



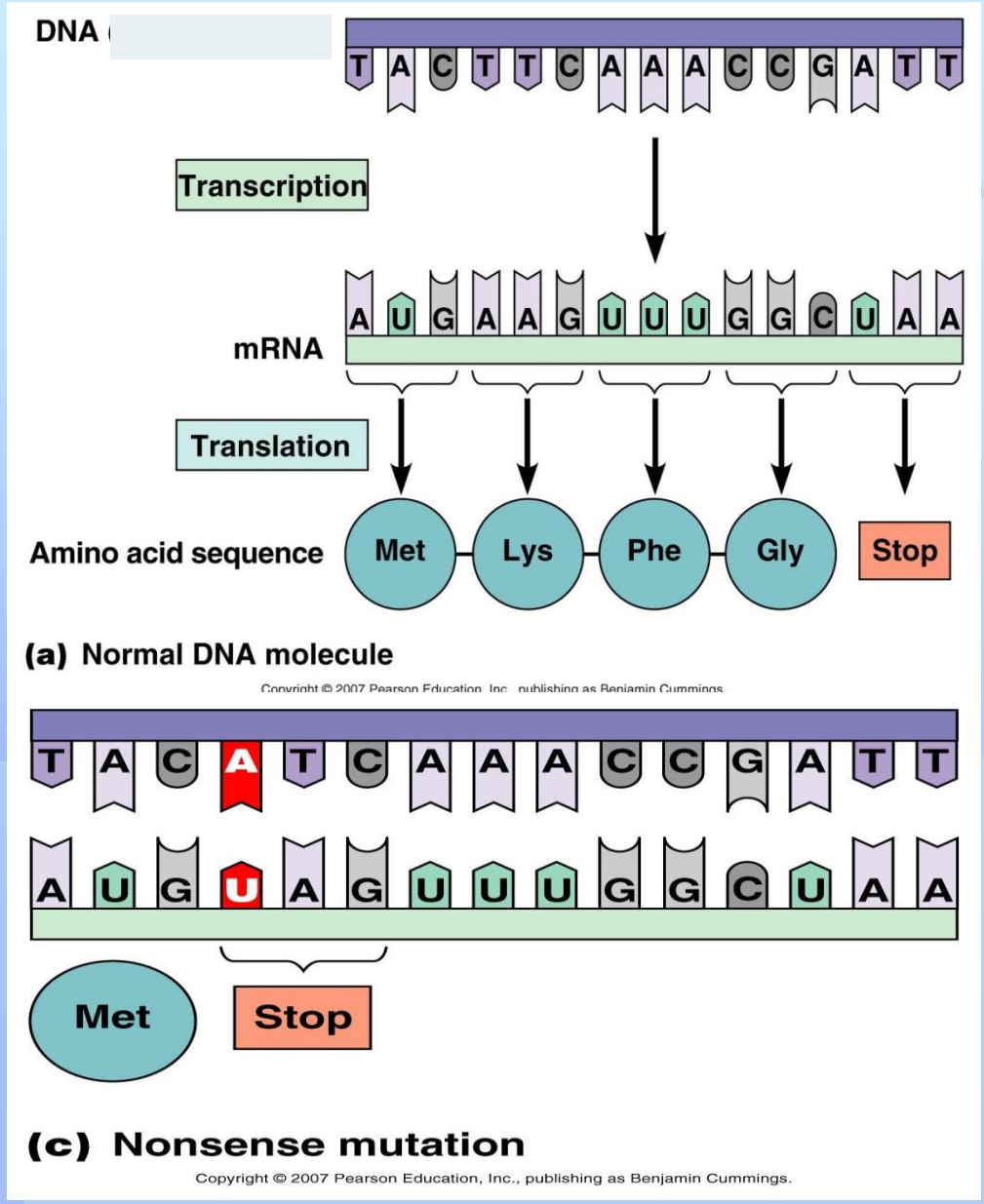
Tyrosine (Tyr) Y



Tryptophan (Trp) W



Single Nucleotide Substitution in Coding Sequence



Template strand (complementary to RNA)

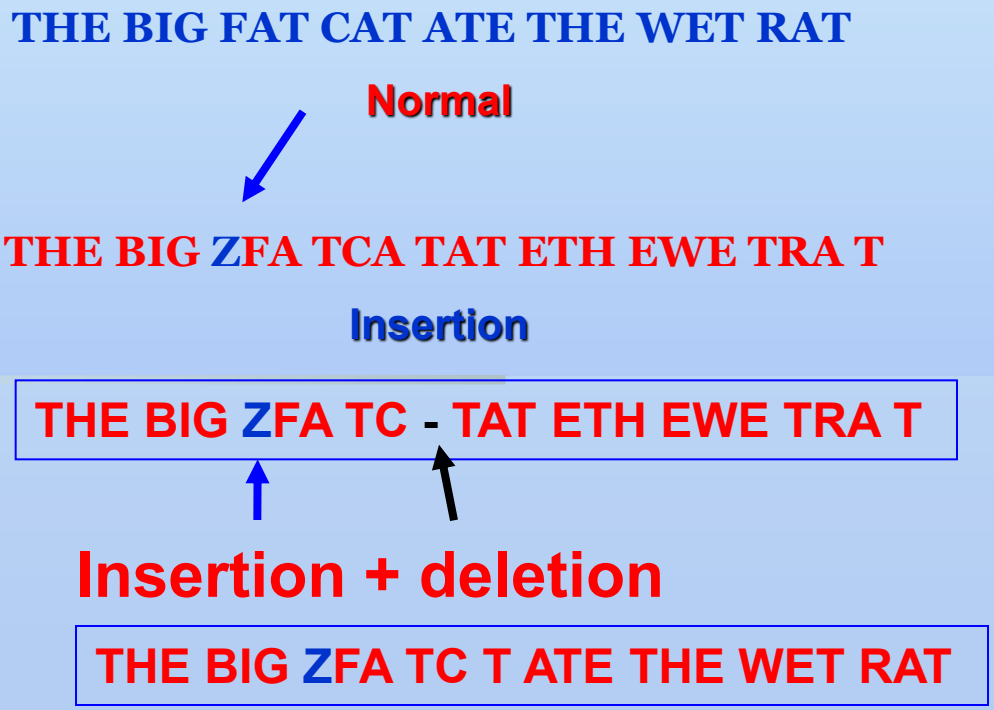
RNA

Protein

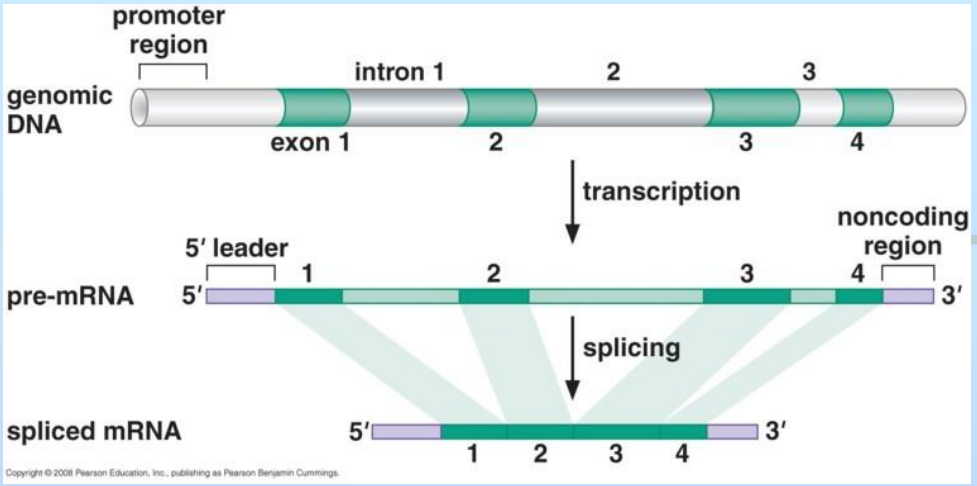
Transition G – A in first base
Stop Codon
Non sense mutation

Genetic Mutation

- **Insertion / Deletion**
  - **A mutation caused by the insertion of at least one extra nucleotide base in a DNA sequence.**



Gene Mutations in Non-coding Sequences



1. A mutation may alter the sequence within a promoter

   Up promoter mutations make the promoter more similar to the consensus sequence

   - They may increase the rate of transcription

   Down promoter mutations make the promoter less similar to the consensus sequence

   - They may decrease the rate of transcription

2. A mutation can also alter splice junctions in eukaryotes

Polymorphism

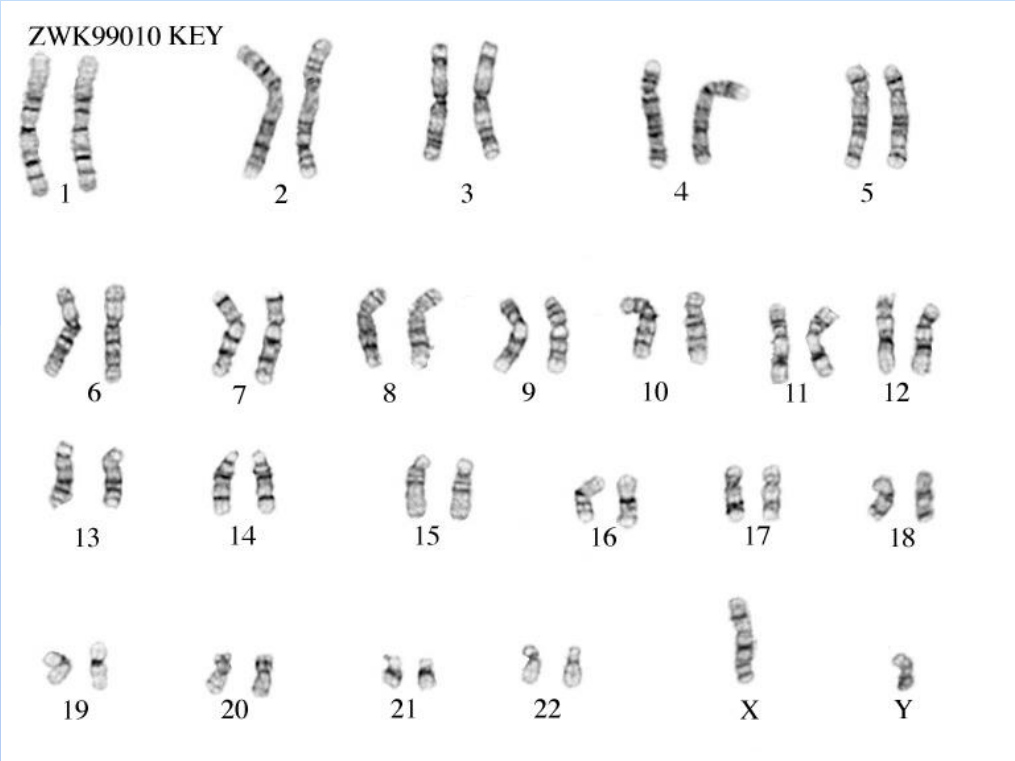
- Several alleles for a single gene
- Polymorphism is an EMPIRICAL CONCEPT that derives from mutations which spread in the population under investigation
- By definition a polymorphic allele > 1% frequency in the population
- It is population-specific
- Multi-polymorphism is the norm

HOW MANY POSSIBLE ALLELES FOR A GENE?

BIOLOGY REMINDERS BOX: chromosomes, kariotypes and meiosis

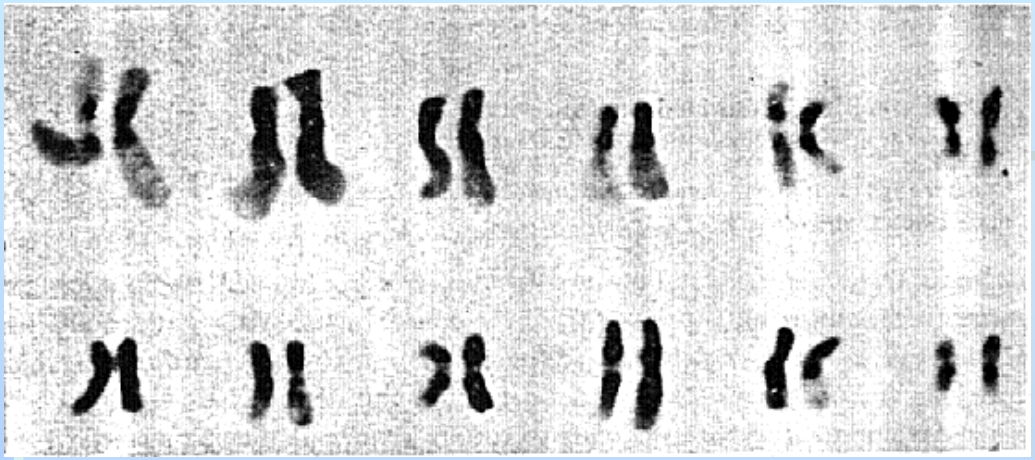
Karyotype of Eukaryotic Cells: the first description of genomes

In each diploid eukaryotic cell there are two HOMOLOGOUS chromosomes for each type of chromosome, *e.g.* in humans 23 pairs = 46 chromosome
Chromosome content in a human diploid cell is 2n = 46
Also 2x, where x = number of pairs of homologue

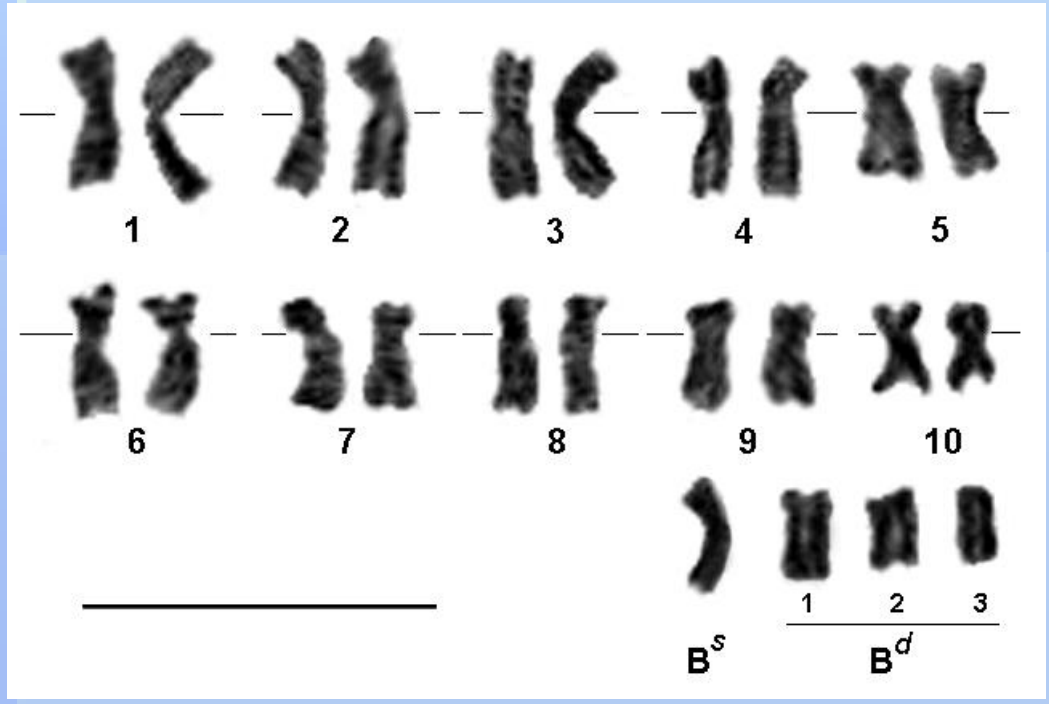


Human karyotype

Plant Karyotypes



Rice karyotype

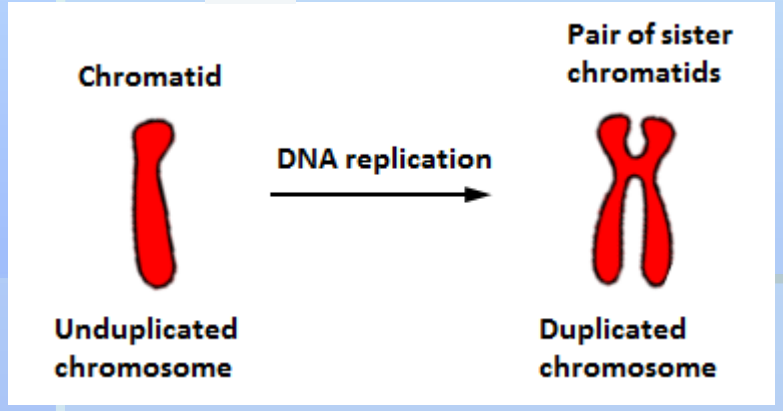


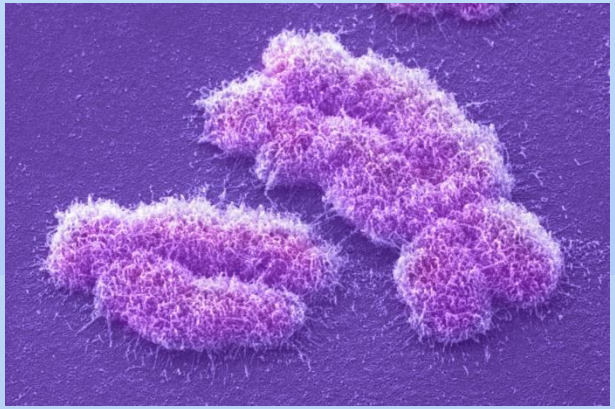
Maize karyotype

MITOSIS

Cellular Process by which from one single eukaryotic cell two daughter cells are produced that are equal among each other and equal to their mother cell

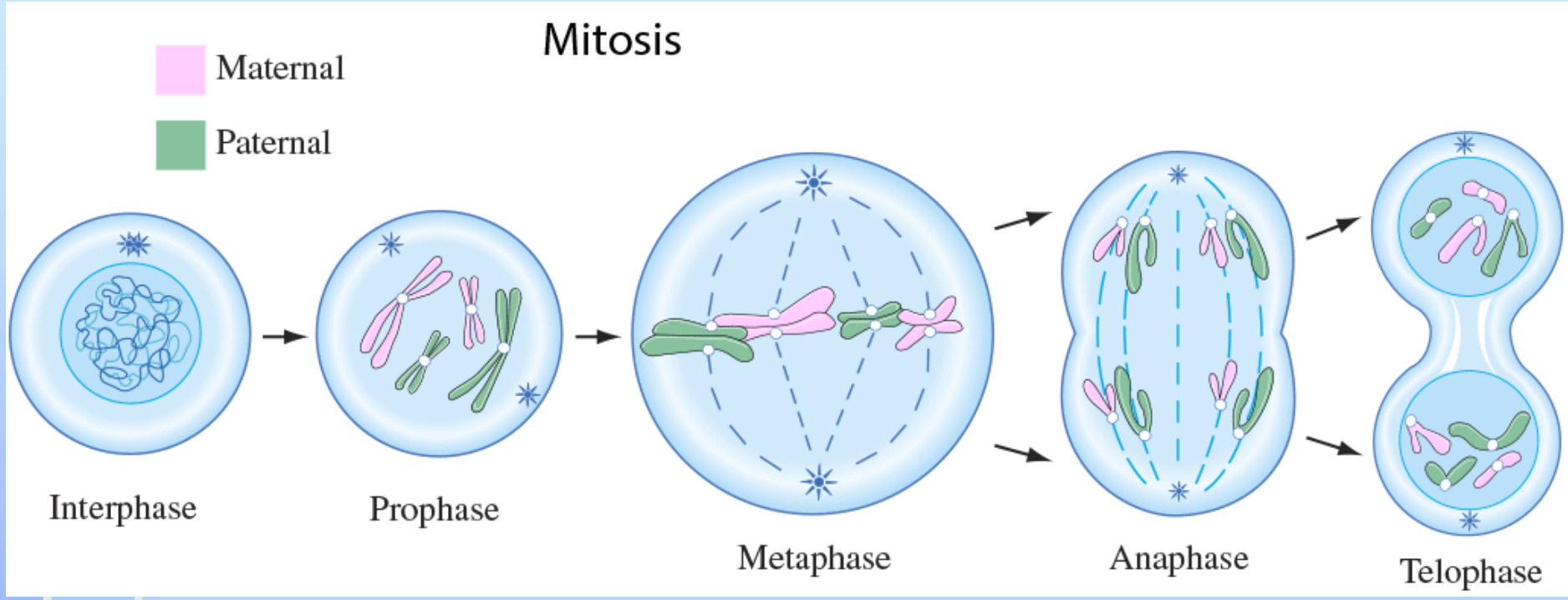
a. One replication of cellular chromosomes → doubling of chromosomes = 4n





b. Only one cell division, therefore from one 4n cell → two 2n cells

MITOSIS



NO new genetic variation is expected

MEIOSIS

Cellular Process by which from one single eukaryotic cell four daughter cells are produced whose chromosome content is half that of the mother cell

IF the cell is diploid (2n) the results are four aploid (n) cells or gametes

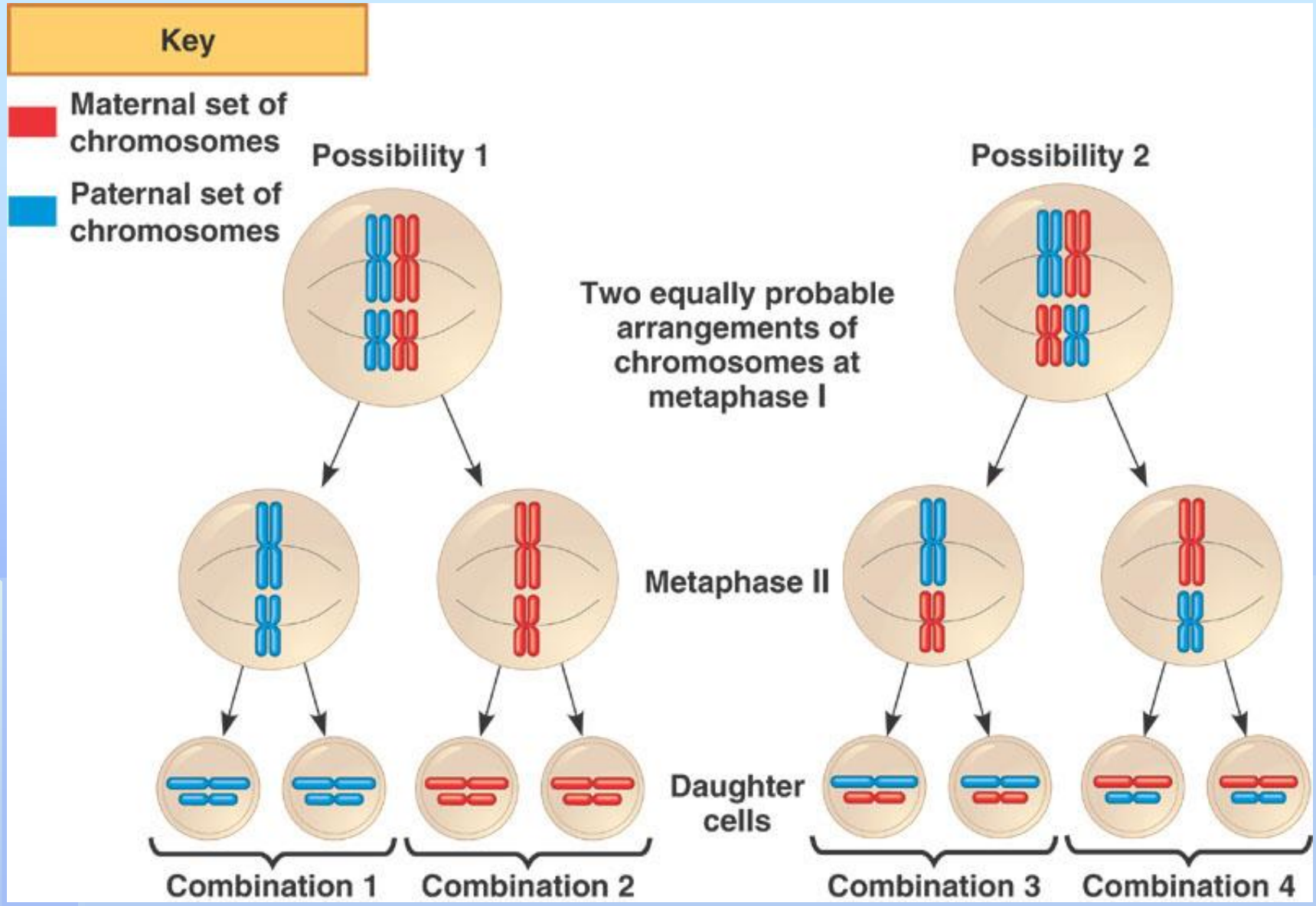
a. One chrosomal replication → one 4n cell

b. Two cellular divisions → four n cells

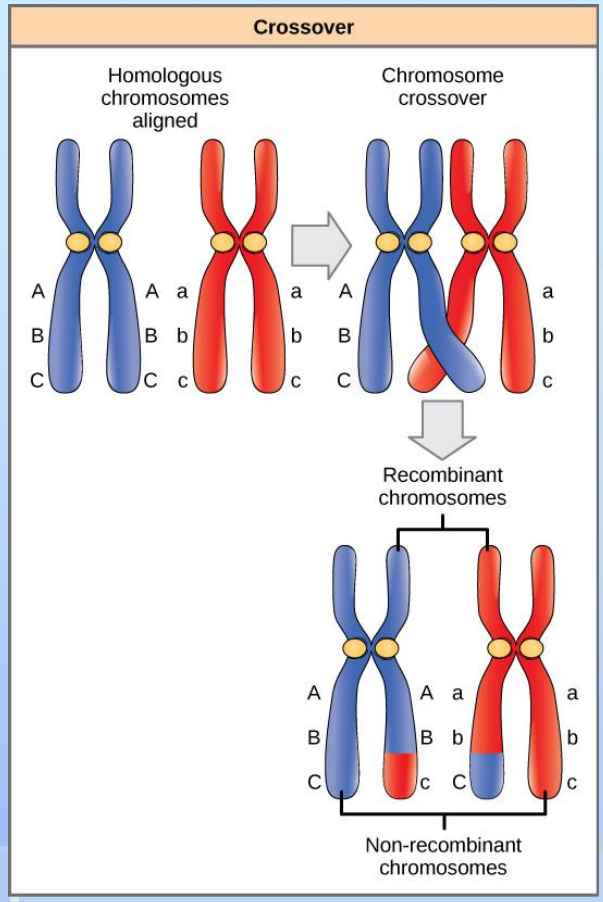
c. The first division is characterized by pairing of homologous chromosomes and determines the halfing of the number of chromosone

d. Second meiotic division is basically a mitosis

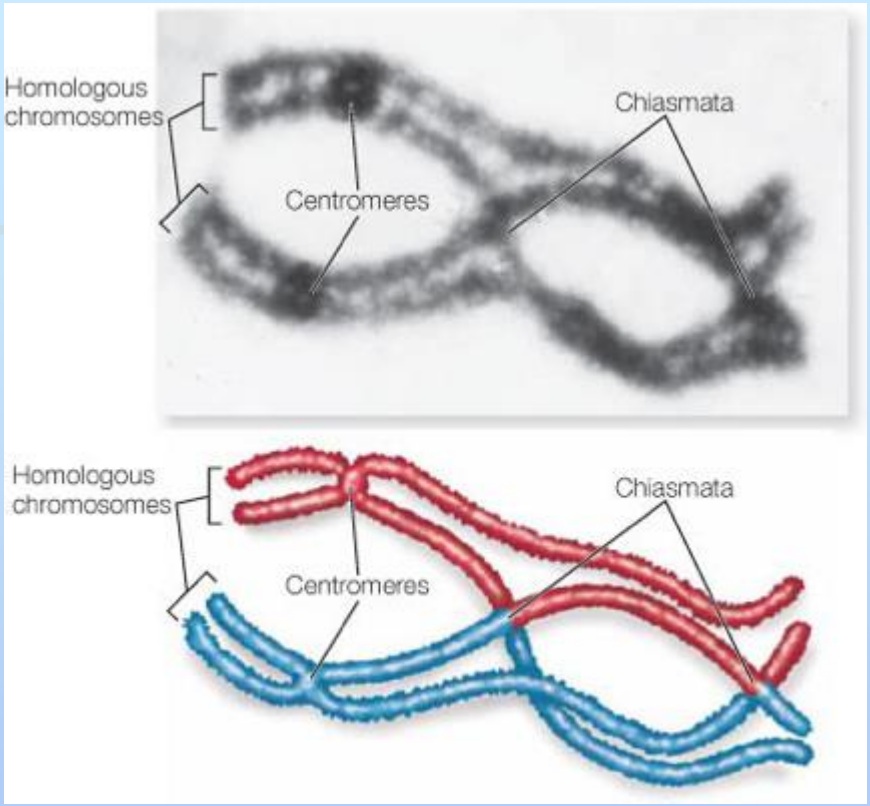
Meiosis and Independent Segregation



Chromosome pairing



Crossing over



Crossing over occurs between two chromatides at a time. C.O. could involve multiple events involving all 4 chromatides

MEIOSIS IN THE PRESENCE OF GENETIC VARIATION

Segregation: the distribution of allelic sequences between daughter cells at meiosis

Crossing over: the exchange of genetic material between homologous crhomosome

Recombinant: a gamete that contains a combination of allelles that is different from the combination inherited by parents

RECOMBINATION IS PRODUCES BY BOTH SEGREGATION AND CROSSING OVER

Thomas Hunt Morgan Demonstrated the Chromosome Theory of Inheritance

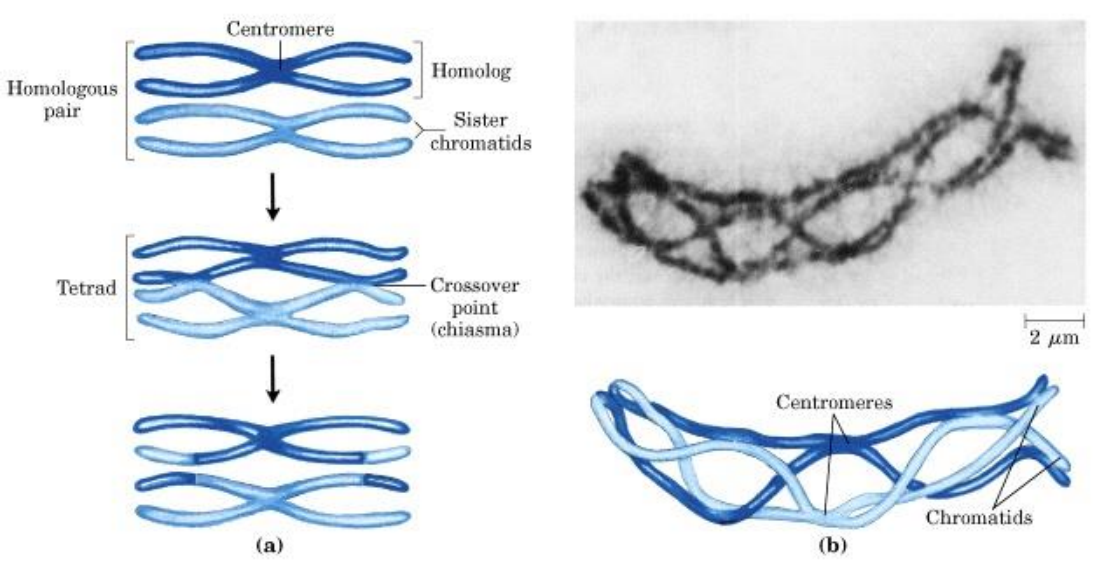
1866 – 1945
American biologist and geneticist
Winner of Nobel prize in Physiology in 1933

The first to associate a specific gene with a specific chromosome

- Like Mendel, Morgan made an insightful choice of his experimental model, *Drosophila melanogaster*
- Fruit flies are prolific breeders and have a generation time of two weeks
- Fruit flies have three pairs of autosomes and a pair of sex chromosomes n = 4

PROBLEM: THERE ARE MORE GENES THAN CHROMOSOMES

GENES and CHROMOSOMES

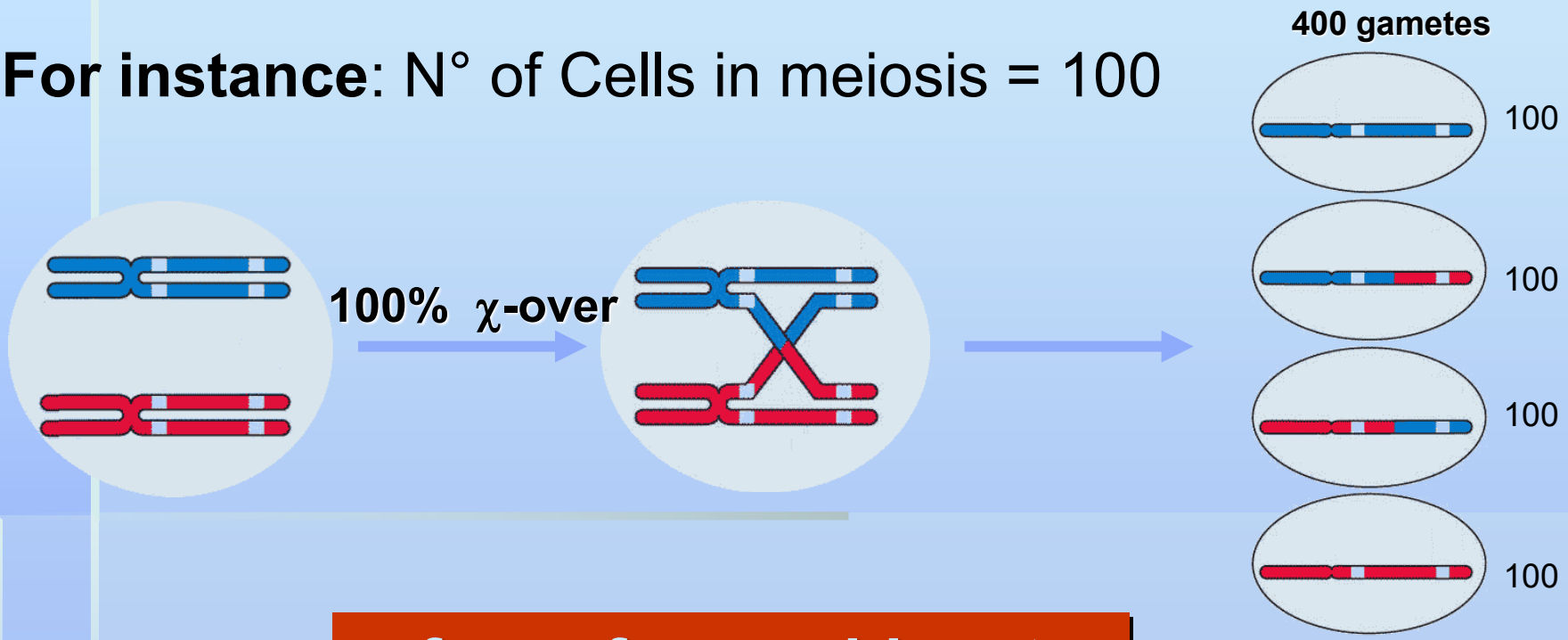


Crossing over occurs between two chromatides at a time. C.O. could involve multiple events involving all 4 chromatides

FREQUENCY of CROSS-OVER and FREQUENCY of RECOMBINATION

Maximum value of RECOMBINANTS is 50%

For instance: N° of Cells in meiosis = 100



freq of recombinant gametes = 50%

Alfred Sturtevant and the FREQUENCY of RECOMBINATION to estimate genetic distances between genes

Alfred Henry Sturtevant

Photo courtesy of Cold Spring Harbor Laboratory Archives.

American geneticist
1891 – 1970
PhD student of Thomas Morgan

Proposed his hypothesis in 1913
When he was still a PhD student

Alfred Sturtevant's Background knowledge

- Maximum percentage of RECOMBINANTS due to segregation is 50% (Mendel's law)
- If two genes are on the same chromosome and there is always a c.o. between them, the maximum percentage of recombinant is 50%

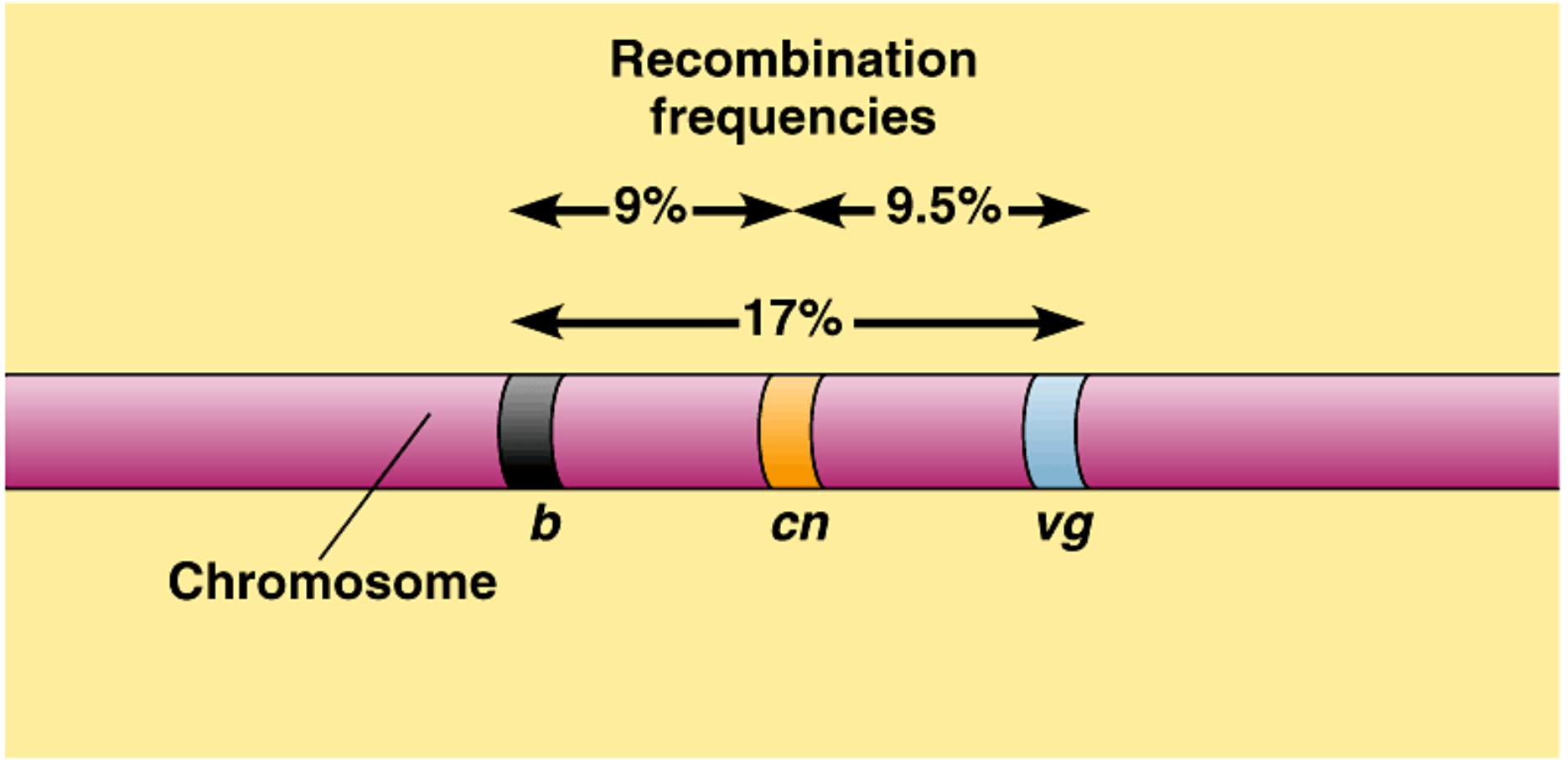
no diffrence from independent segregation

Alfred Sturtevant's Hypothesis of linkage between genes and estimate of genetic distance

Hypothesis: If c.o. is a **random event**, its frquency would depend on the distance between genes. Therefore the Frequency of recombination can be an **estimate of the genetic distance between the two genes**

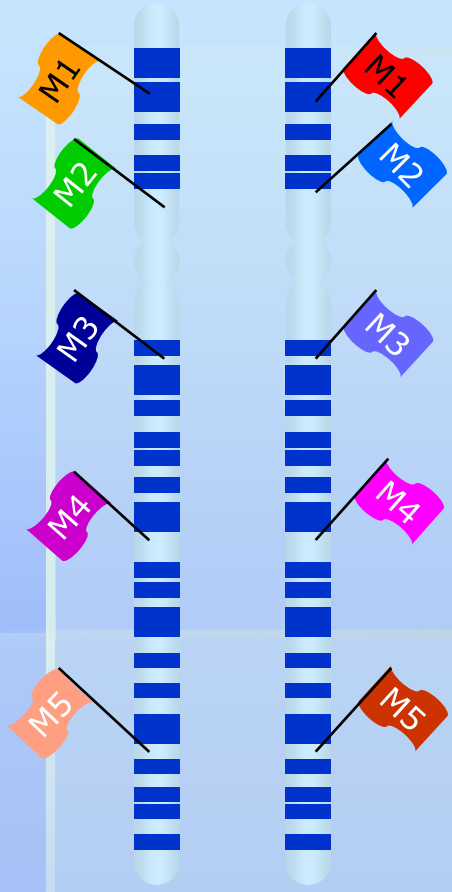
NOTE THAT THE FREQUENCY OF RECOMBINATION CAN BE ESTIMATED ONLY IF THERE ARE 2 DIFFRENT ALLELES FOR EACH OF THE GENES CONSIDERED

Using Frequencies of Recombination to Construct Genetic Maps



Integrating molecular biology into genetic mapping

Genetic Markers



Genetic Marker

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

- Tool → not the objective
- Marker → indirect approach

Genetic markers

TOOLS FOR GENETIC ANALYSIS

morphological

- Classical Mendelian traits

biochemicals

- isoenzymes
- structural proteine (endosperm)

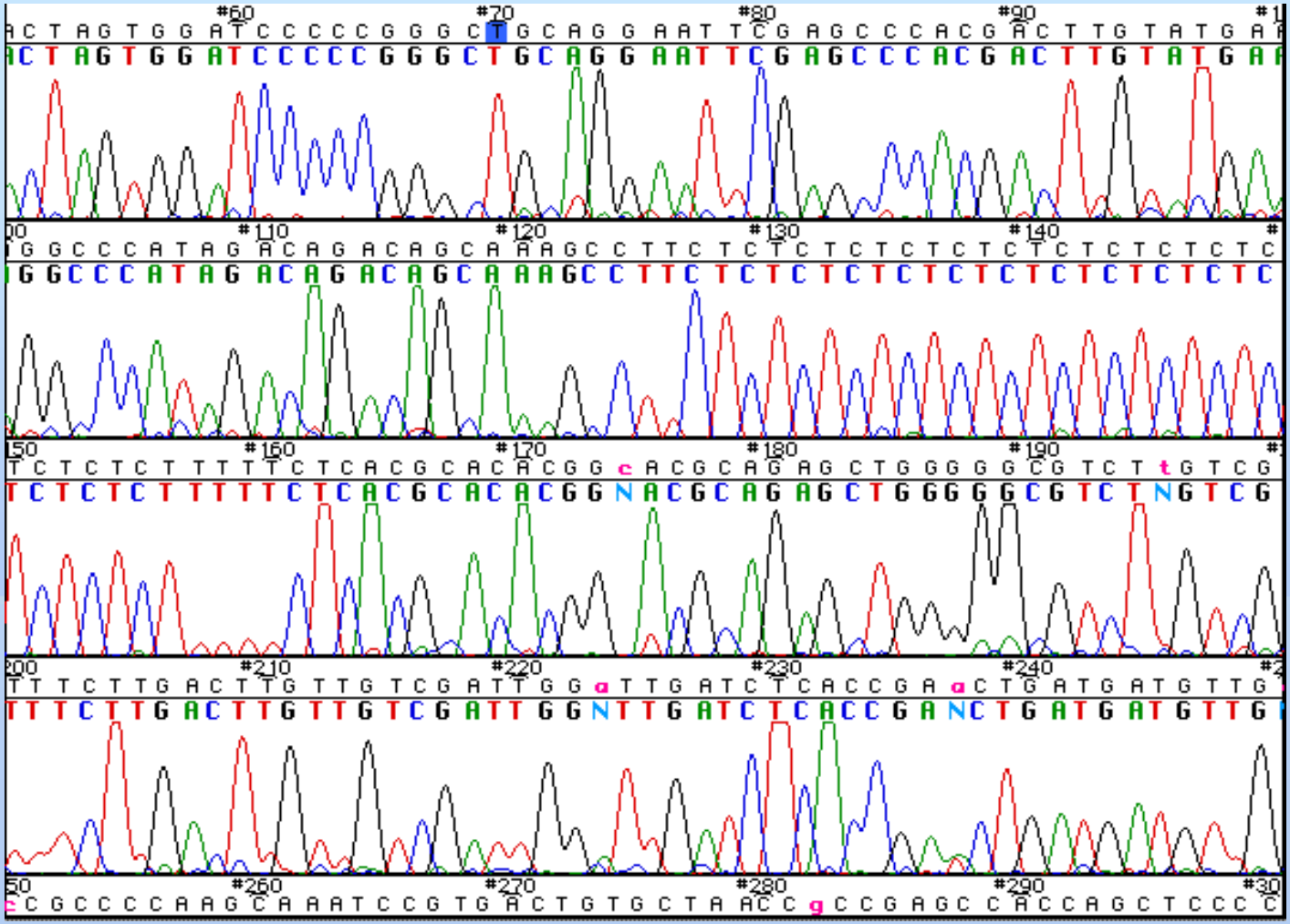
molecular

- based on DNA (RFLP, VNTR, RAPD, microsatellites, SSCP, AFLP, Single Nucleotide Polymorphisms

Ideal GENETIC Marker: Characteristics

- Mendelian behaviour
- Not influenced by the environment
- Highly polymorphic
- Easy to detect and analyze
- Robust
- Not expensive
- Automation

PCR for SSR Analysis



METHODOLOGICAL BOX

POLYMERASE CHAIN REACTION – PCR –

Kary Mullis (1944 -)
American Biochemist
Cetus Corp.
Nobel prize in Chemistry in 1993

Polymerase: DNA polymerase

- DNA polymerase REPLICATES DNA

Chain Reaction: DNA polymerase

- The product of a reaction is used to amplify the products of each reaction

It is the PERFECT MEHODOLOGY to amplify regions containing SSRs and identify GENETIC VARIATION AT THE DNA LEVEL

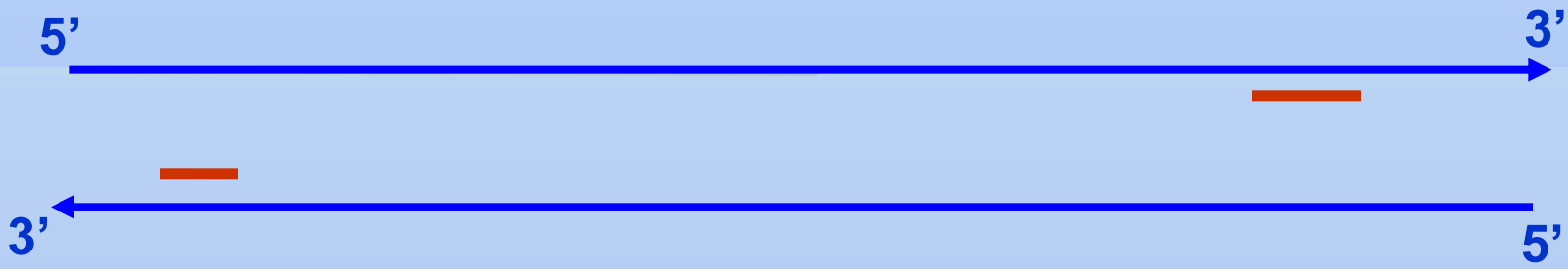
Properties of DNA polymearse

It needs a pre-existing DNA to duplicate

- Cannot assemble a new strand from components
- Called template DNA

It can only extend an existing piece of DNA

- Called primers

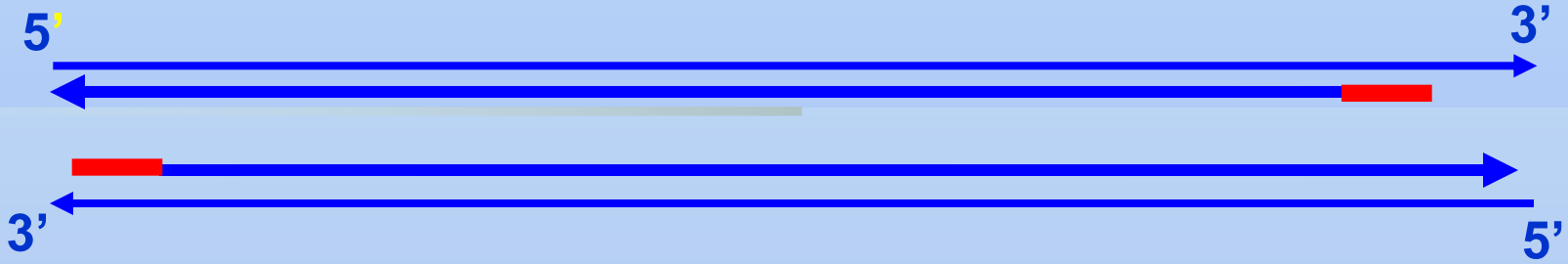


Properties of DNA polymearse

DNA strands are anti-parallel

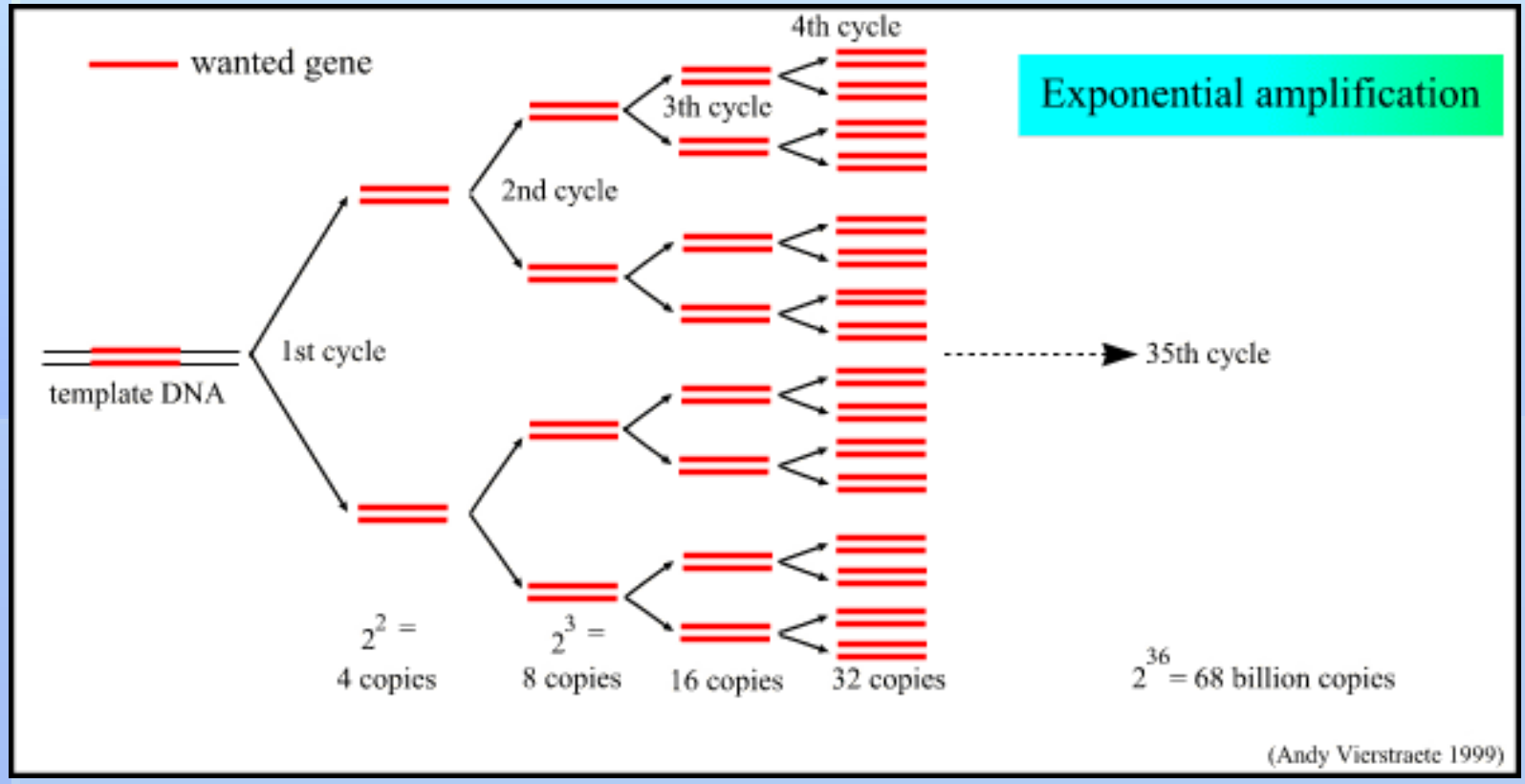
- One strand goes in 5’ → 3’
- The complementary strand is opposite

DNA polymerase always moves in one direction (from 5’ → 3’)



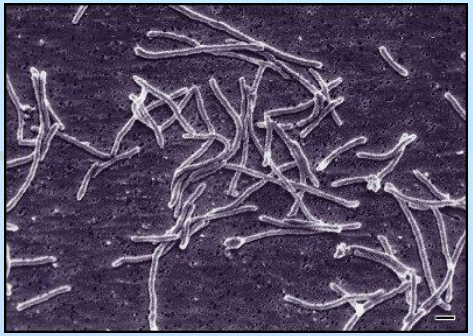
PCR Cycling

- The newly generated DNA strands serve as template DNA for the next cycle
- PCR is very sensitive
- Widely used



Taq DNA polymerase

- Derived from *Thermus aquaticus* thermophylus bacterium
- Heat stable DNA polymerase
- 1000 nt /sec at 72°C
- No proof reading activity



Lower Geiser at Yellowstone

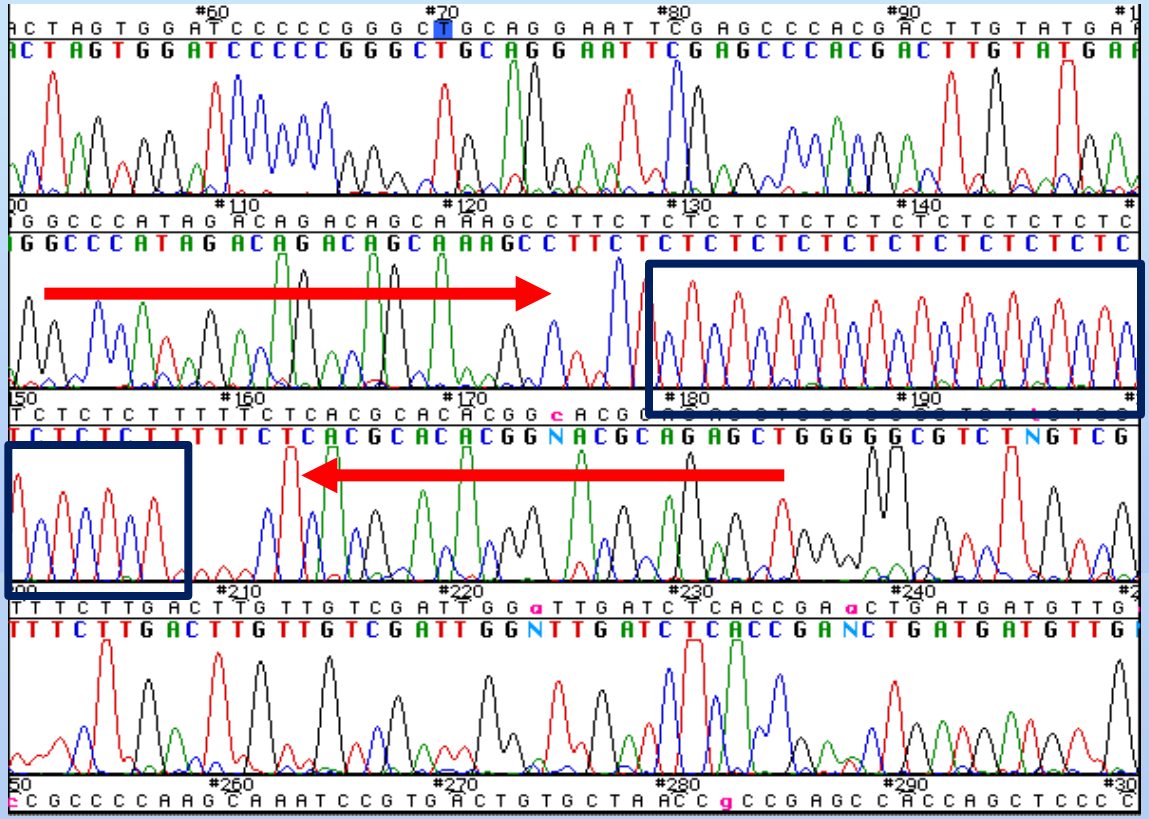
Thermal Cycling

A PCR machine controls temperature

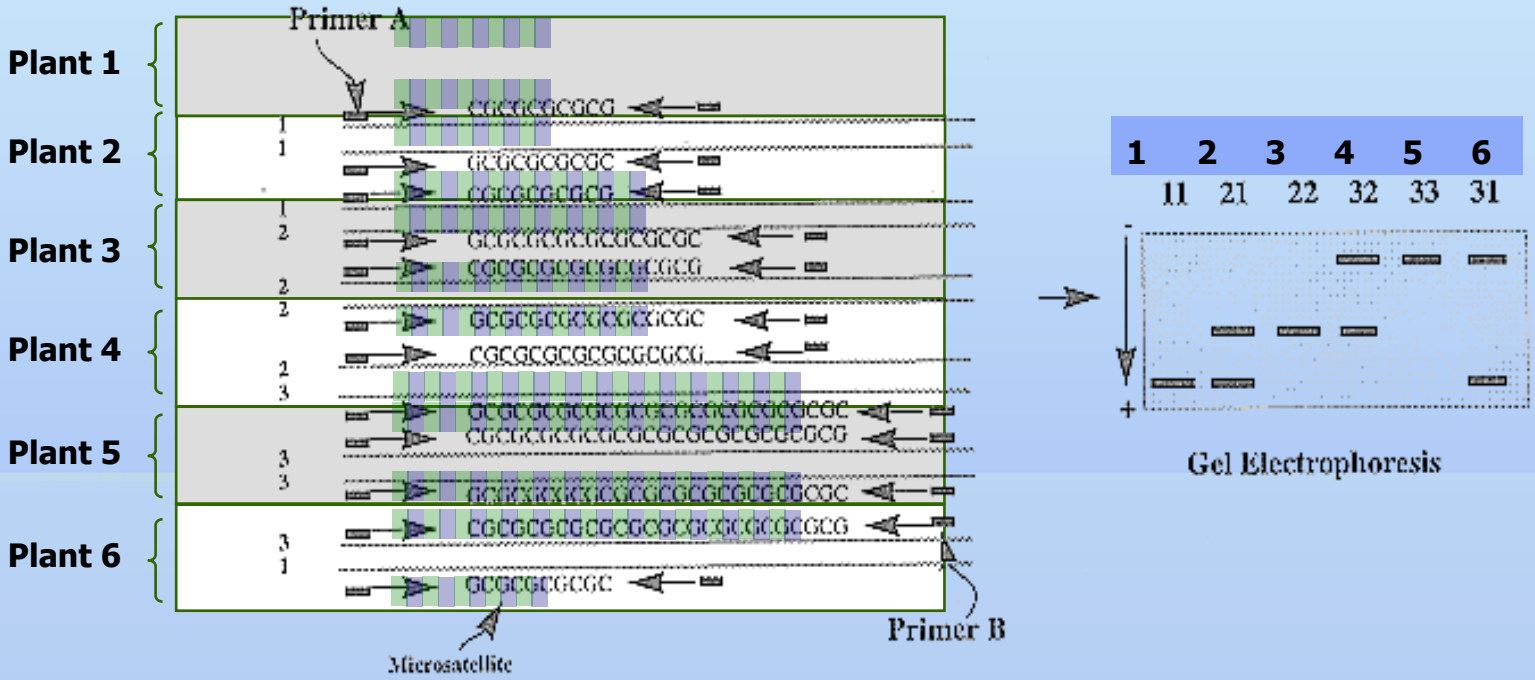
Typical PCR go through three steps

- Denaturation
- Annealing
- Extension

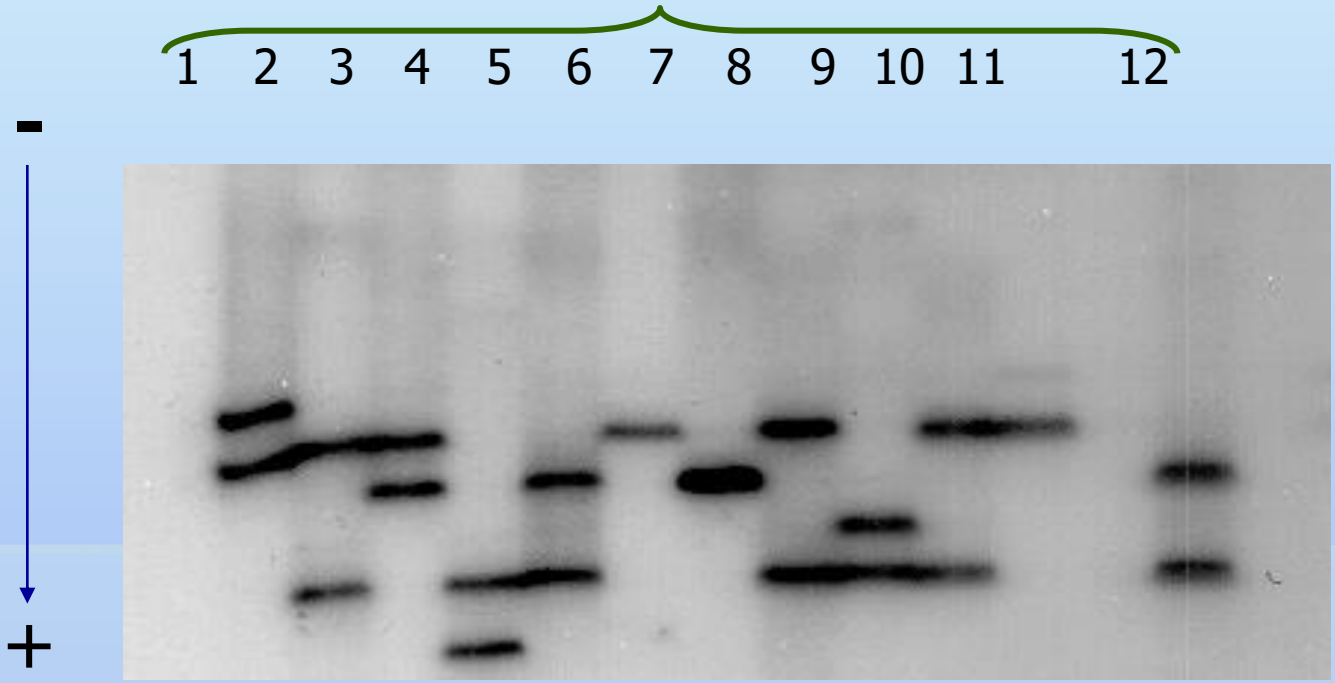
Simple Sequence Repeat - SSR – (Microsatellites)



Simple Sequence Repeats - SSR – (Microsatellites)

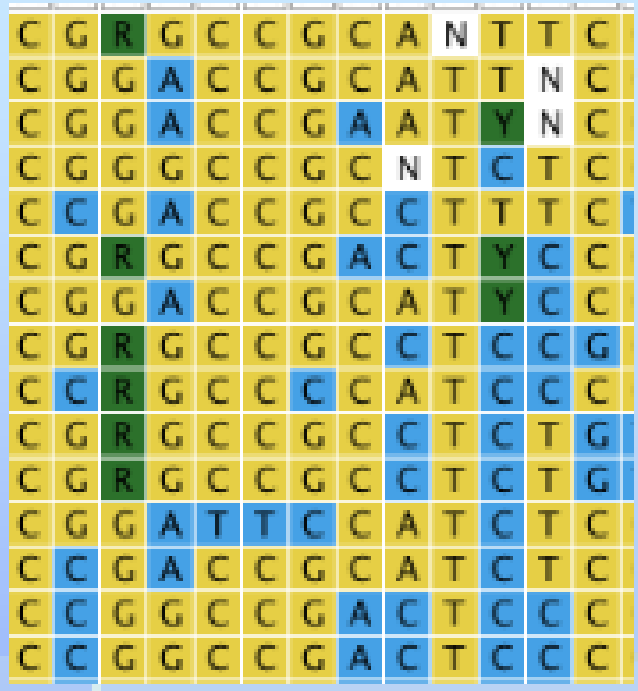


SSR Analysis of different individuals from a natural population



Separation of different alleles by PAGE

Single Nucleotide Polymorphism - SNP -

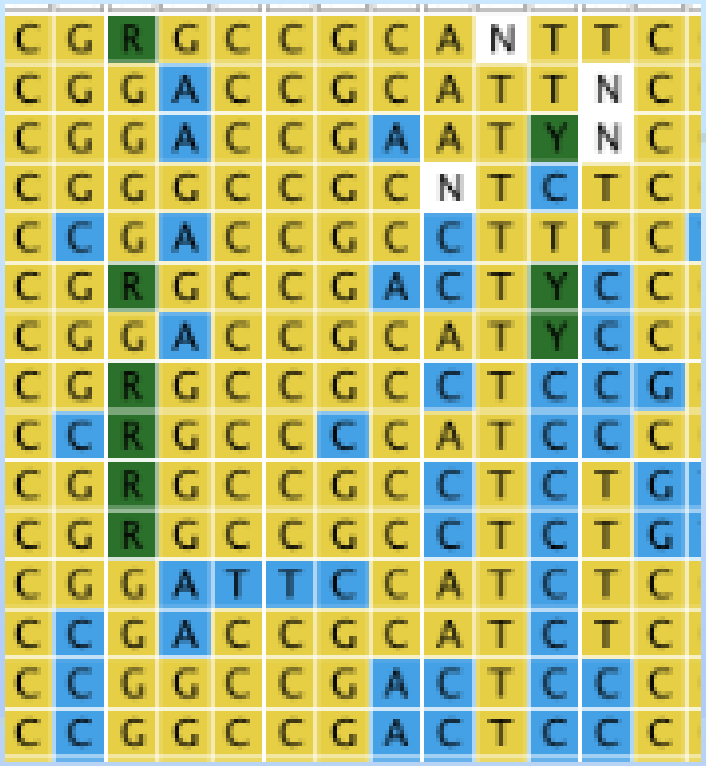


…C C **A** T T G A C…
…G G **T** A A C T G…

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

- Identified by sequencing
- Their frequency depends on several factors

Allele Identification is NOT the Bottleneck Anymore



In the genomics era the challenge is to assign these sequences a meaning

Molecular Markers Applications

- **Population Analysis**
  - Population Genetics
  - Taxonomy and Evolution
- **Mapping**
  - Linkage maps
  - Mapping of single genes
  - Mapping of complex traits (QTL)
  - Dignostics
  - Marker assisted breeding
  - Gene Isolation
- **Fingerprinting**
  - Cultivar Identification
  - Forensic
  - Germplasm characterization

GENETIC MAP

Map

- It defines linear relationships among *loci*
- Produced by genetic analyses
- Distance between genes is measured as frequency of recombination

IT IS THE RESULT OF GENETIC EXPERIMENTS

1 centi-Morgan (cM) = 1 recombinant / 100 meiotic prodoucts

Polymorphism

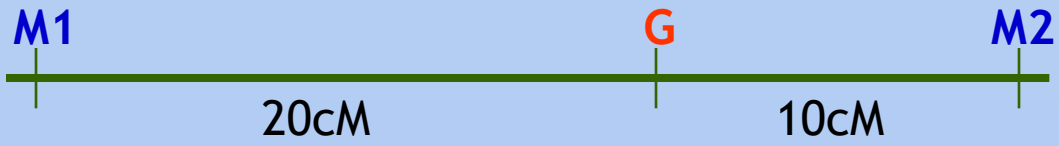
- Allelic Variants for a specific gene

Molecular Polymorphism

- Differences at the DNA level

Molecular Marker

- Genetic *locus* that identifies a position on the genetic map



Relevance of a Genetic Map

- Integration with known mutants
  - ⇨ FUNCTION
- Quantitative Traits
  - ⇨ IMAPPING PHENOTYPES
- Genomic Organization
  - ⇨ COMPARATIVE STUDIES

Genetic Maps are an application of FORWARD GENETICS

STEPS

1. Identification of polymorphism
   → Molecular markers generation
2. Selection of parental genotypes and breeding
3. Production of a segregating population
   → F2 – Backcross – etc.
4. Genotyping of single individuals in the population
   → Alleles at polymorphic *loci*
5. Analysis of segregation data
   → Genetic Map

Types of Mapping Populations

F_2

- All heterozigous *loci* segregate in a single meiosis

Back-cross = F_1 x P_x

- Only the alleles of the NON recurrent parent segregate

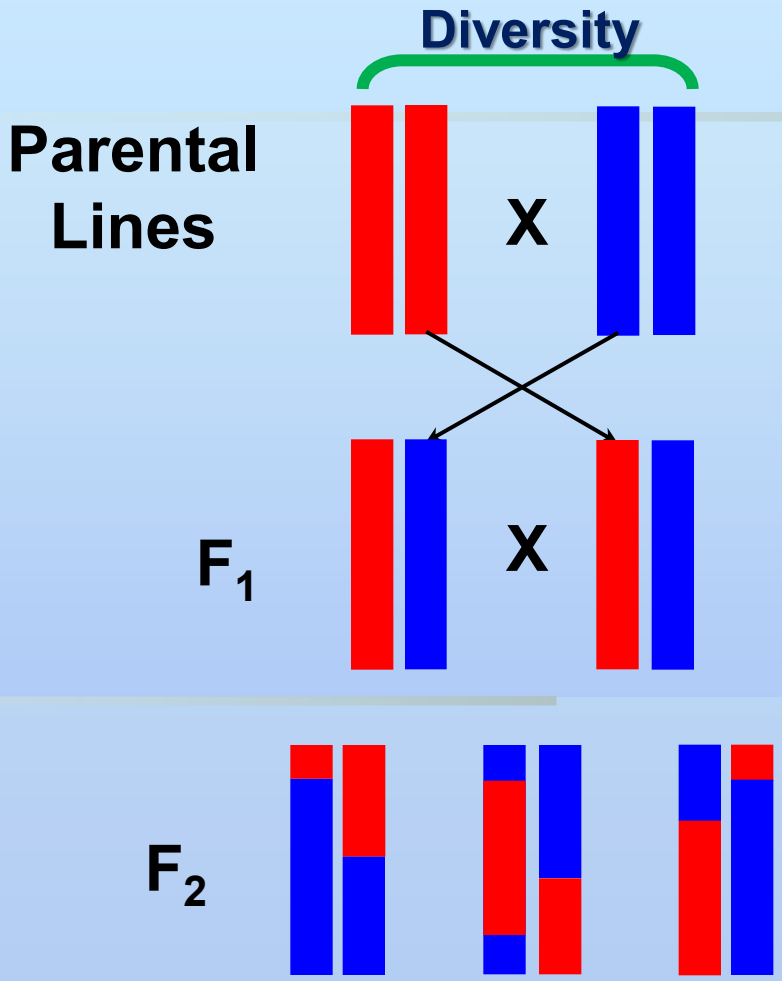
Ricombinant Inbred Lines (RIL)

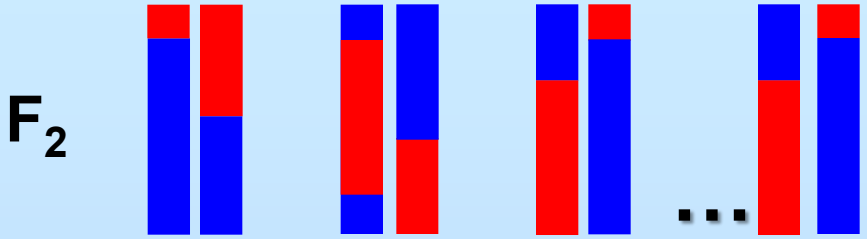
- All heterozigous *loci* segregate in multiple meiosis

Pseudo back-cross

- F1 x C where C is NOT homozygous

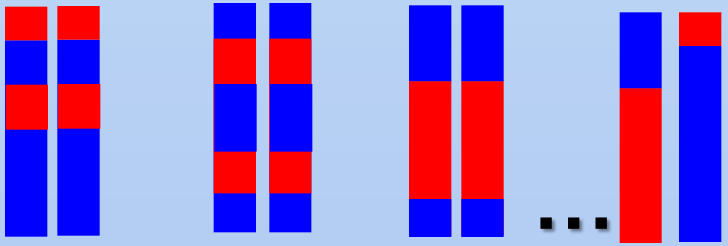
First steps in RILs production



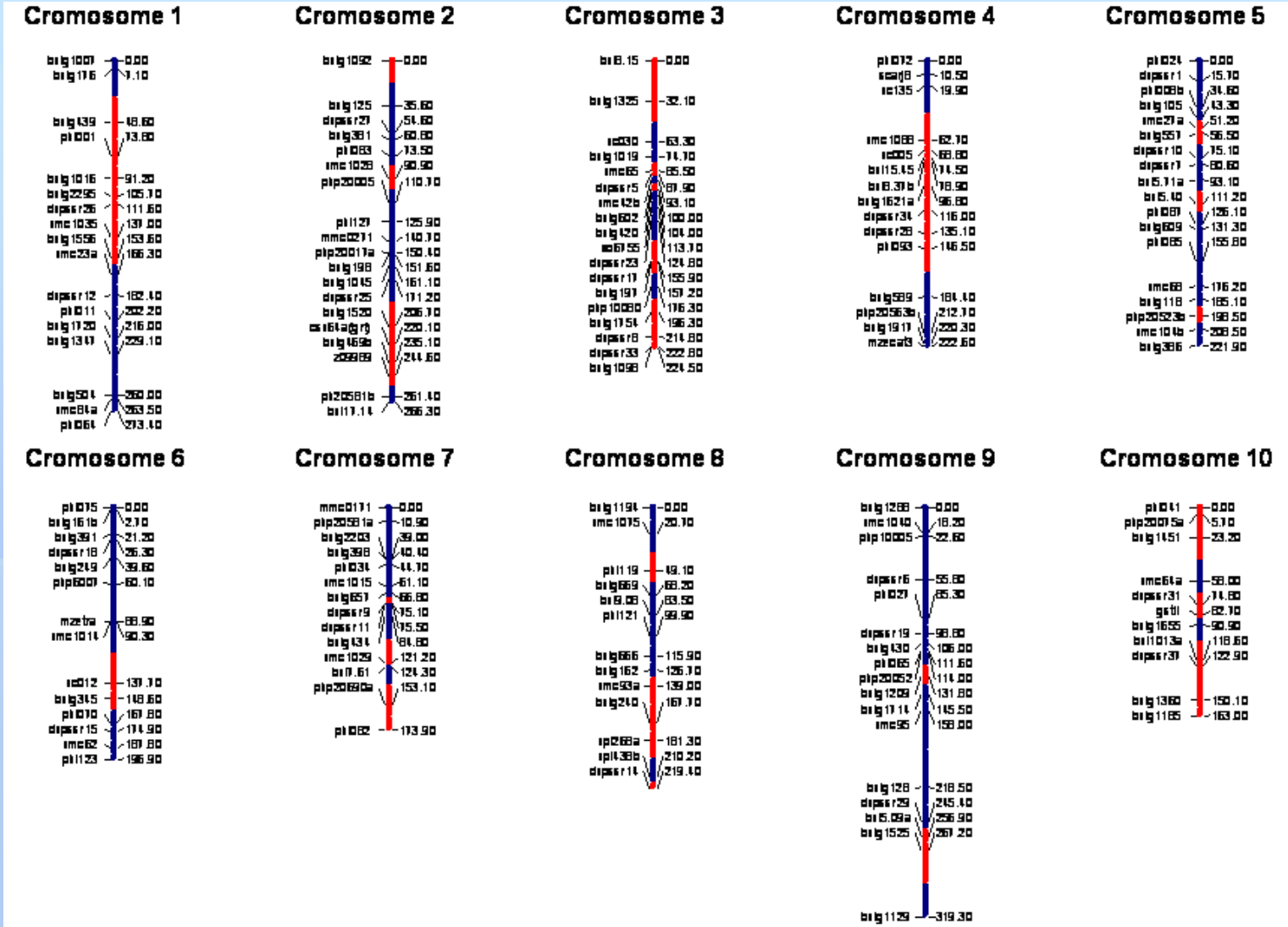


Several Cycles of Inbreeding

RIL families are immortalized genotypes: recombination blocks are fixed



Mapping Pop: B73 x H99 Recombinant Inbred Lines



Genetic Mapping in the Genomic Era

- Efficient way to MAP phenotypes
- Tool to better understand the relationship between genotype and phenotype
- Tool to derive indication on the function of DNA sequences (genes; regulatory elements)
- Based on the linkage between DNA polymorphisms and phenotypic variation:

  1) STATISTICAL EVIDENCE

  2) 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.
- Therefore it defines a fixed position on the genome
- 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 chromosomal positions.
- Appropriate materials (crossing or natural populations) are required
- Molecular characterization of the population
- Consistent and reproducible phenotypic data

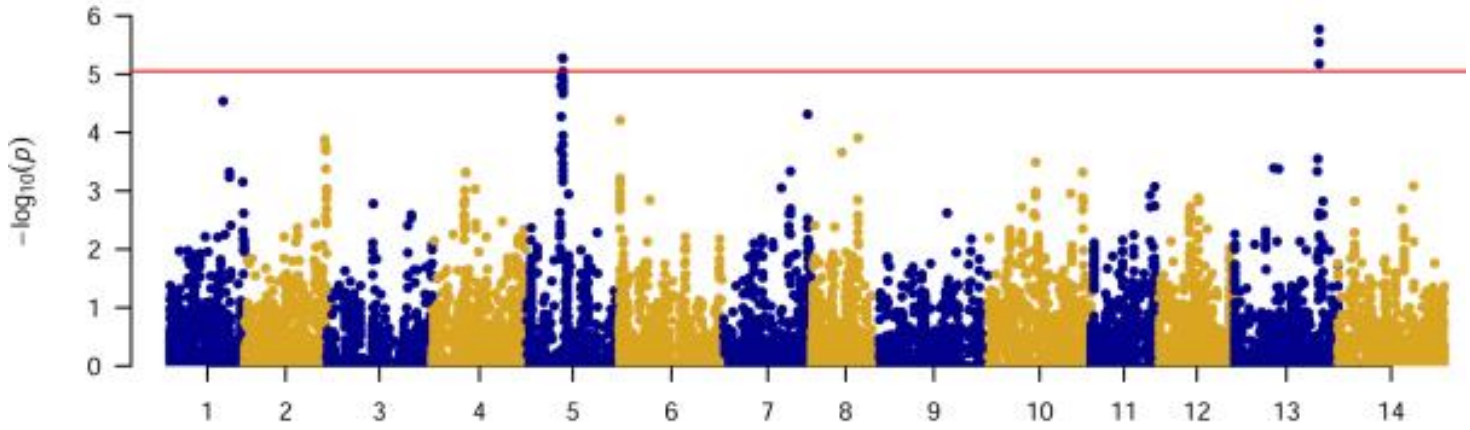
Genomics and Mapping of Traits

Genomic Data

C	G	R	G	C	C	G	C	A	N	T	T	C
C	G	G	A	C	C	G	C	A	T	T	N	C
C	G	G	A	C	C	G	A	A	T	Y	N	C
C	G	G	G	C	C	G	C	N	T	C	T	C
C	C	G	A	C	C	G	C	C	T	T	T	C
C	G	R	G	C	C	G	A	C	T	Y	C	C
C	G	G	A	C	C	G	C	A	T	Y	C	C
C	G	R	G	C	C	G	C	C	T	C	C	G
C	C	R	G	C	C	C	C	A	T	C	C	C
C	G	R	G	C	C	G	C	C	T	C	T	G
C	G	R	G	C	C	G	C	C	T	C	T	G
C	G	G	A	T	T	C	C	A	T	C	T	C
C	C	G	A	C	C	G	C	A	T	C	T	C
C	C	G	G	C	C	G	A	C	T	C	C	C
C	C	G	G	C	C	G	A	C	T	C	C	C

Statistical Analysis

Phenotypic Data



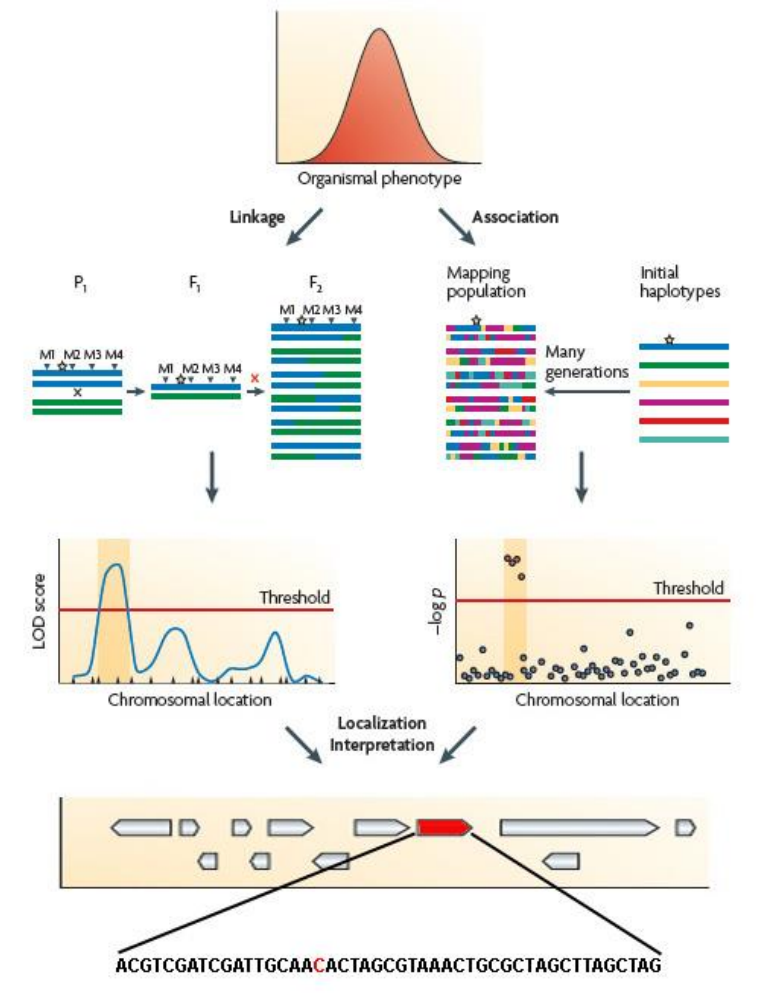
Identification of chromosomal regions controlling the variation of trait(s) First step toward gene identification and isolation

QTL are mapped through two different methods

Linkage mapping through artificial populations

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

Structured pops



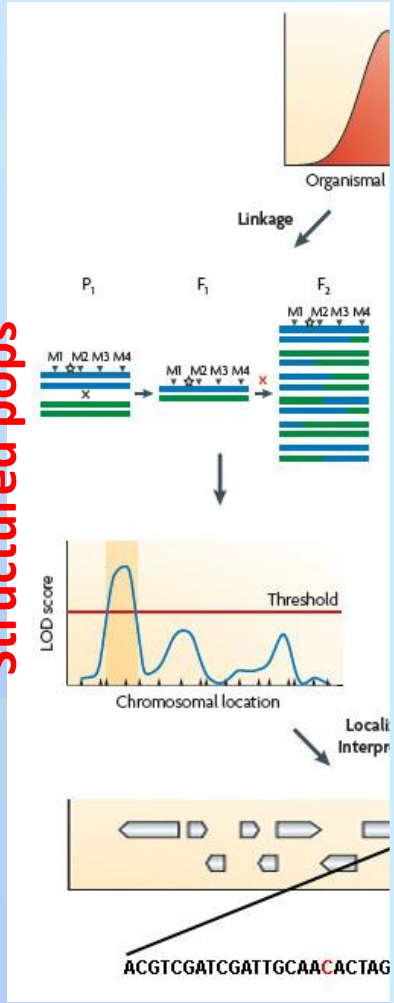
Diversity pan

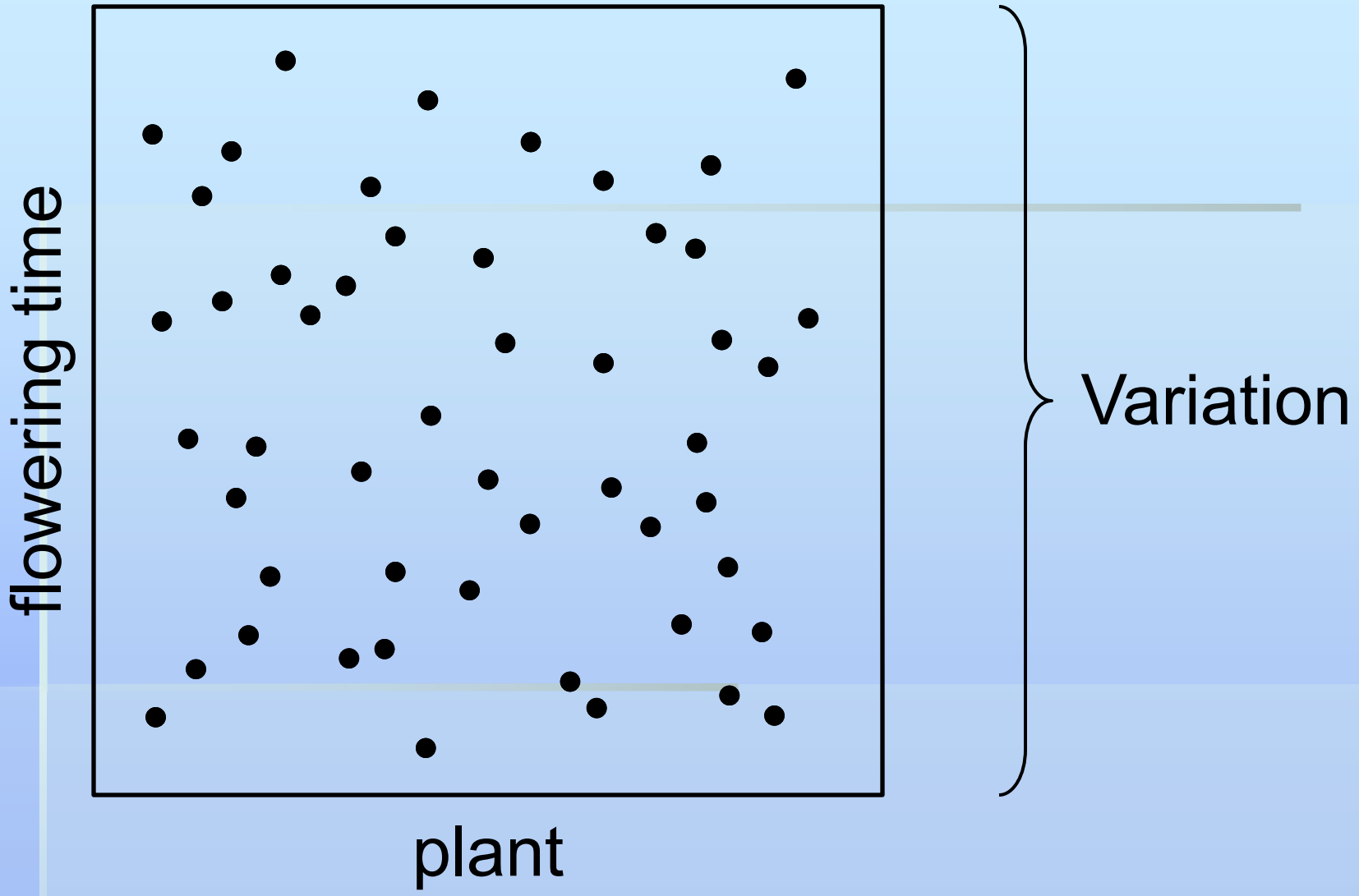
Association (GWAS) through "natural populations"

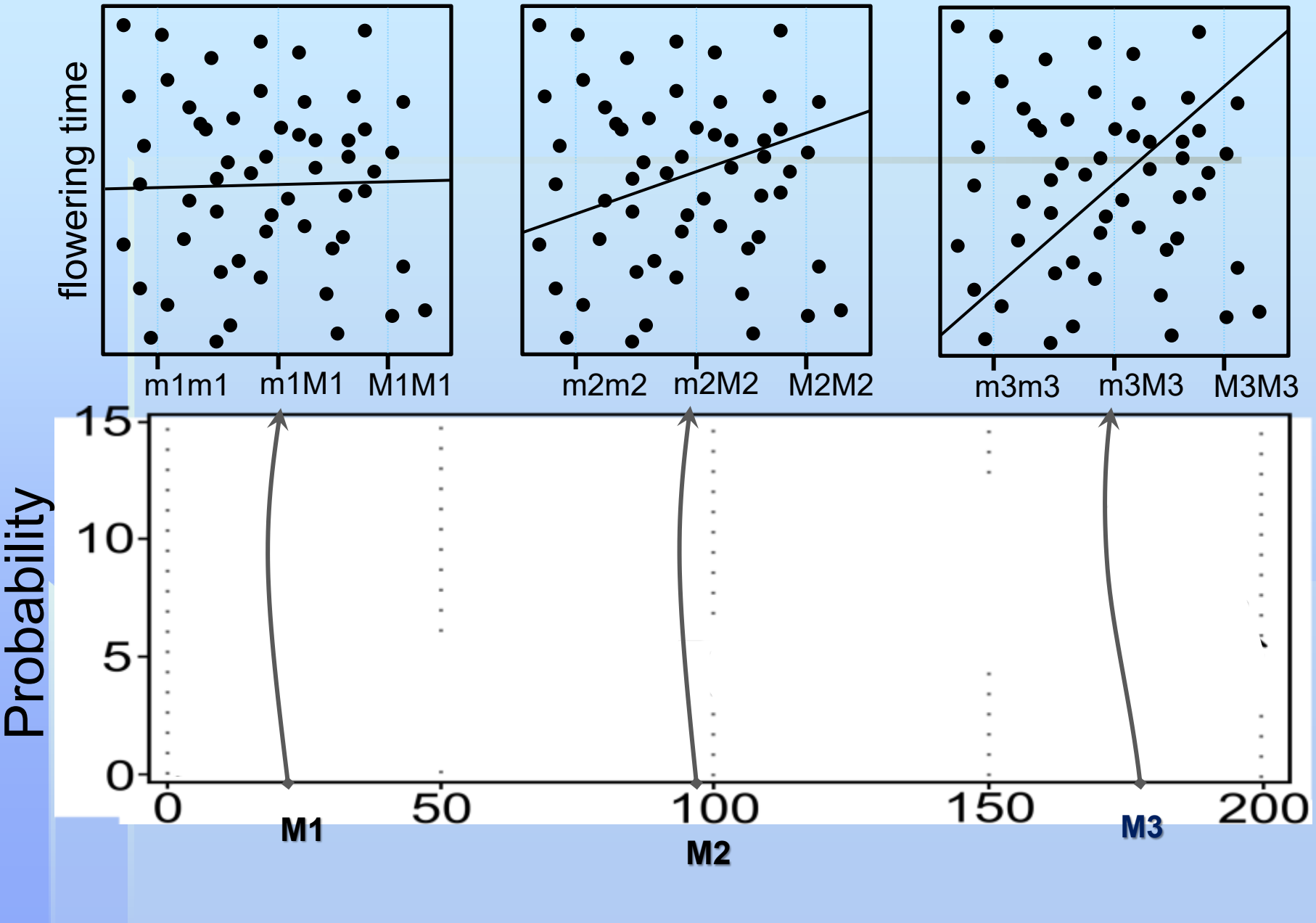
- high density necessary
- hidden structure, LD
- higher false rate
- broad variation
- high definition
- faster, cheaper

A closer look to QTL mapping in structured artificial populations

Structured pops

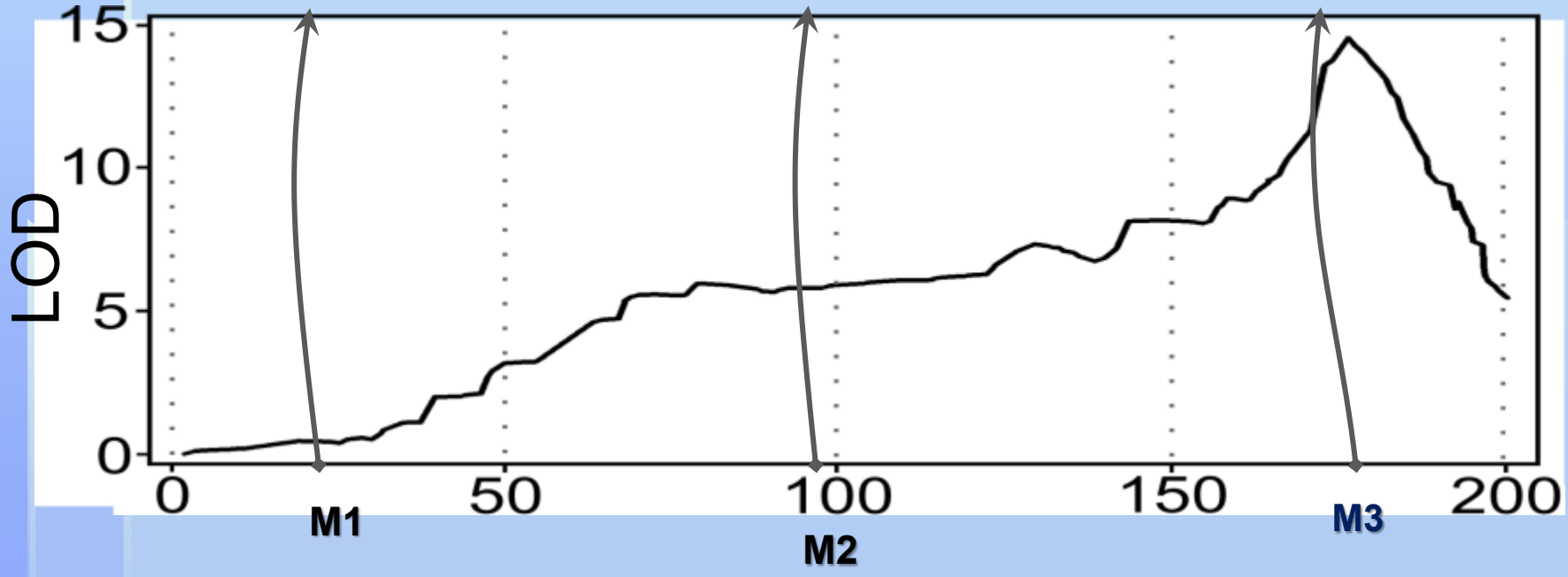



flowering time
plant
Variation


flowering time
m1m1
m1M1
M1M1
m2m2
m2M2
M2M2
m3m3
m3M3
M3M3
Probability
15
10
5
0
0
50
100
150
200
M1
M2
M3

Interval Mapping

- Interval mapping is conceptually similar to linkage analysis to a single marker. However, the marker is replaced by a position within the interval between 2 markers
- The expected effect of a hypothetical QTL mapped in that position is estimated considering the effects of flanking markers



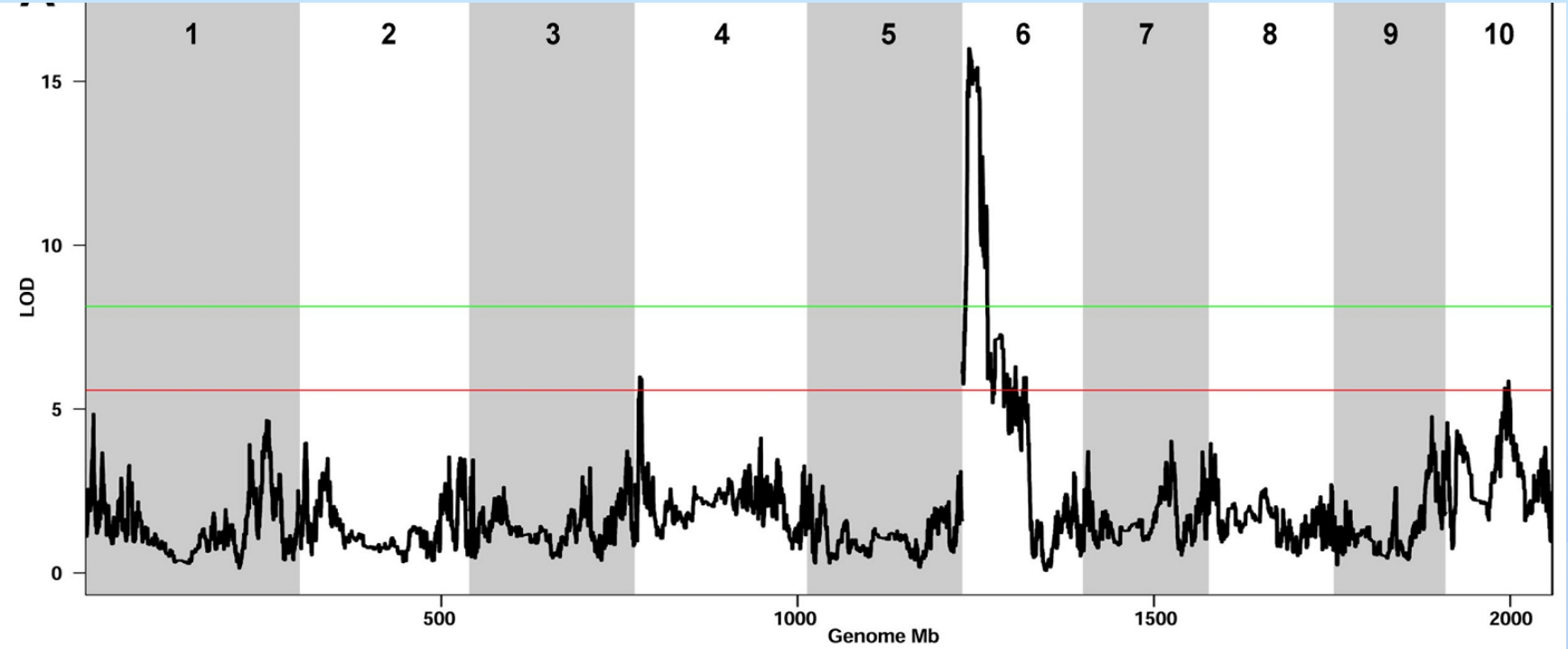
LOD score

LOD = *Logarithm of odds*

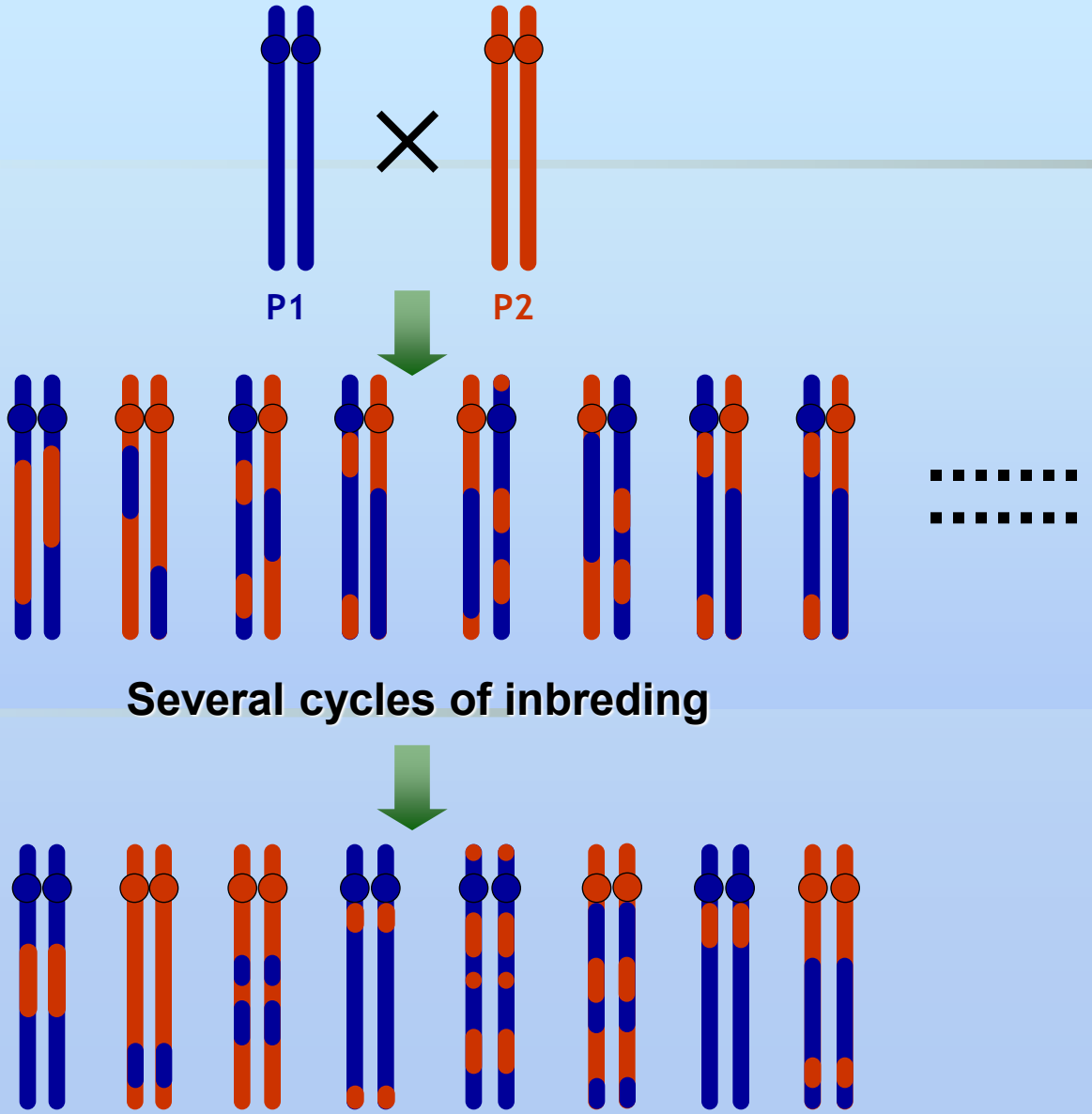
LOD = $$\log\left(\frac{probability\ of\ QTL.present}{probability\ of\ QTL.NOTpresent}\right)$$

LOD = 3 : 1000 HIGHER the probability of a QTL

One Example: QTL for Yield



Recombinant Inbred Lines (RIL)

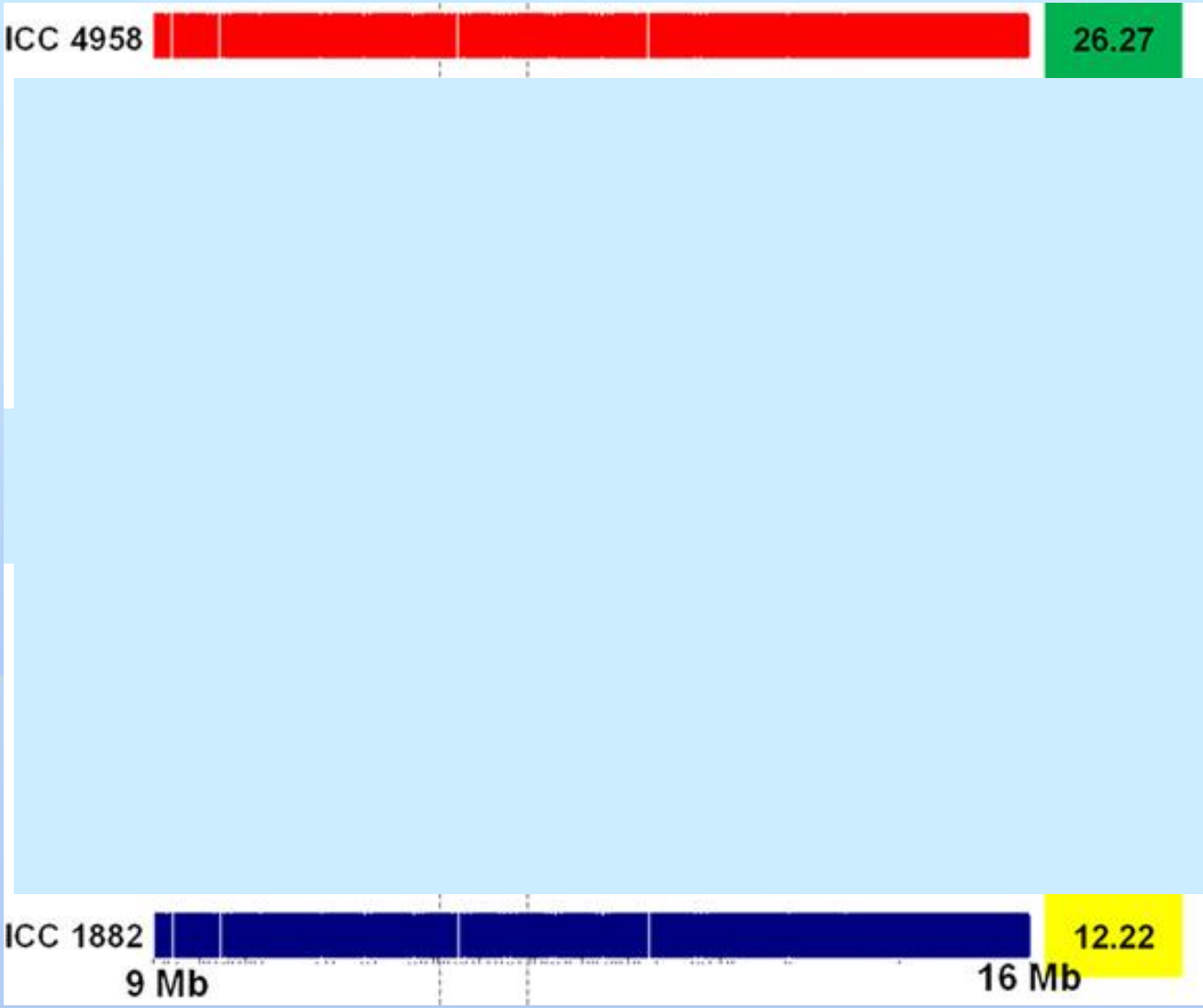


2

chromosome 7

TGW

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

BOX on CRISPER - CAS9
The control of genetic variation

The genetic mantra – no genetic variation no genetics - and the obsession to control genetic variability

The control of genetic variability has always been the Holy Grail of genetics since Mendel and the geneticists who have striven to develop knowledge and methodologies to understand it, control it, and exploit it.

Indiana Jones and the last crusade (1989)

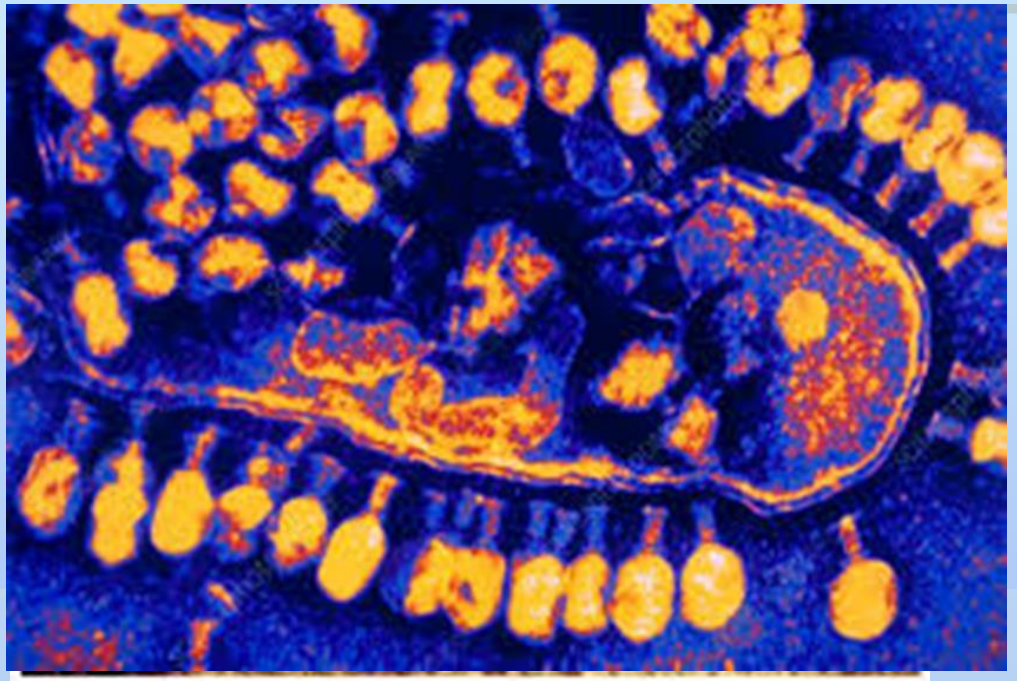
Genome Editing Revolution is Woman
Nobel Prize in Chemistry in 2020

Emanuelle Carpentier
1968 - French

Jennifer Doudna
1964 - American

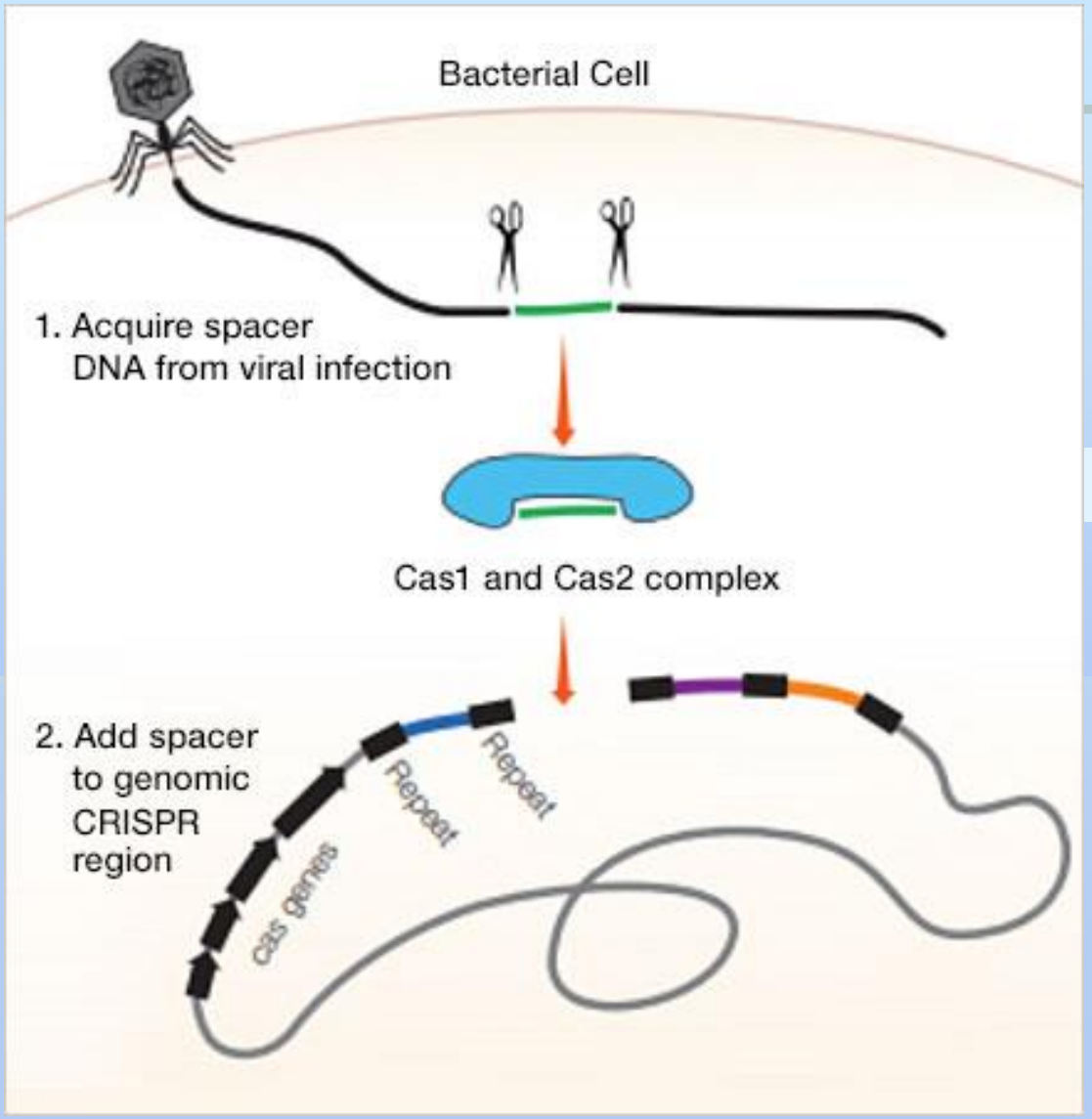
"Charpentier and Doudna discovered the CRISPR/Cas9 genetic scissors. This tool has contributed to many important discoveries in basic research, and plant researchers have been able to develop crops that withstand mould, pests and drought. In medicine, clinical trials of new cancer therapies are underway…. these genetic scissors have taken the life sciences into a new epoch and, in many ways, are bringing the greatest benefit to humankind." (motivation of the Nobel Committee)

Bacterial Cell under viral attack (Batteriophages)



Cellular lysis and virus release

Discovery and Description of the Adaptive Immune Defense Mechanism

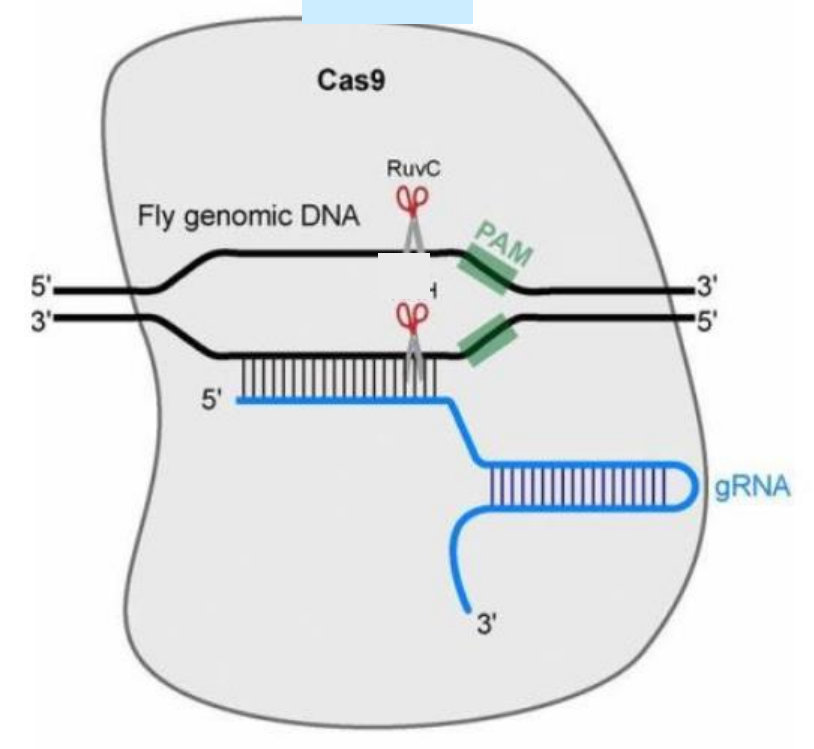


Molecular Players

- **CAS9 protein** is able to cut the DNA double helix in a very precise way → **MOLECULAR SCISSOR**
- **Specific RNA molecules** are produced and form complexes with CAS9 proteins
- **Those RNA molecules** contain a sequence that is complementary to the portion of the bacterial genome acquired by the bacterium that specifically bind the viral DNA, in so doing **they guide CAS9** to the viral DNA called the **RNA Guide**
- **CAS9 cuts the viral DNA** blocking the invasion

THE CUT IS EXTREMELY SPECIFIC and PRECISE

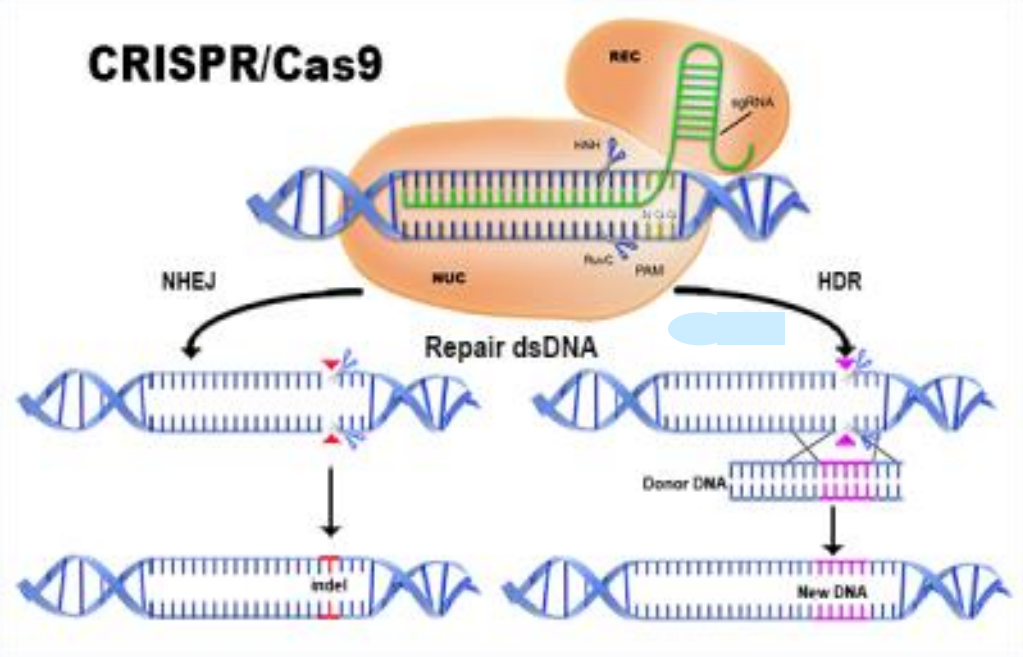
CAS9 – RNA Guide Complex and the Target Site



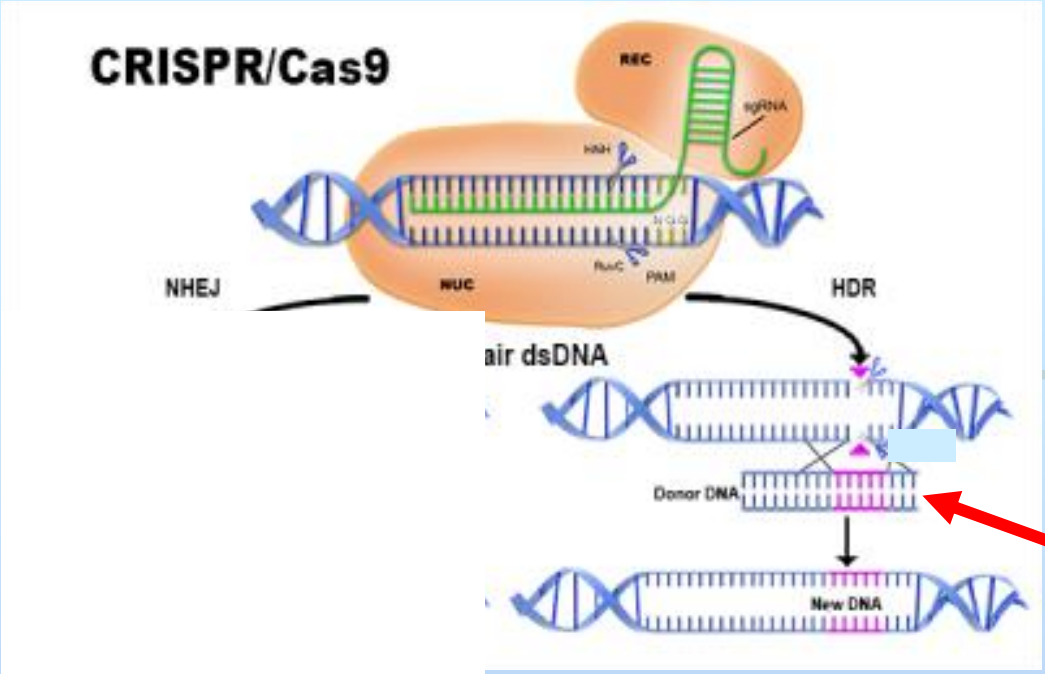
- The Complex is bound to the DNA target
- The double helix is open and accessible to CAS9
- CAS9 cuts the double helix

From Bacteria to higher organisms

1. DNA REPAIR



- Broken DNA is higly reactive
- The break is repaired by cellular DNA repair systems
- It is an SOS repair therefore it is Error prone
- Wrong nucleotides can be inserted and RANDOM MUTATION can be produced in the original site
- The type of mutations are INDISTIGUISHABLE from natural mutations



2. DNA RECOMBINATION

external DNA fragment

- Broken DNA is sealed by integrating a DNA fragment which replaces the broken one
- The mechanism is based on a DNA exchange based on sequence similarity at the end termini of the the broken fragments
- **ADDITION** and **REPLACEMENT** of the original site
- The integrated fragment could be **natural or synthetic**

Crisper - Cas9 limitations

- The principal bottleneck is plant regeneration
- Need to sequence all product for assessing homozygousity and number of mutational events
- Assessment of phenotypic gains
- The most important traits are the results of complex genetic control → MULTIPLE GENES MODIFICATION
- KNOWLEDGE OF THE GENES TO BE MODIFIED AND HOW

This brings us back to QTL mapping

Two major features are required for QTL mapping to be efficient

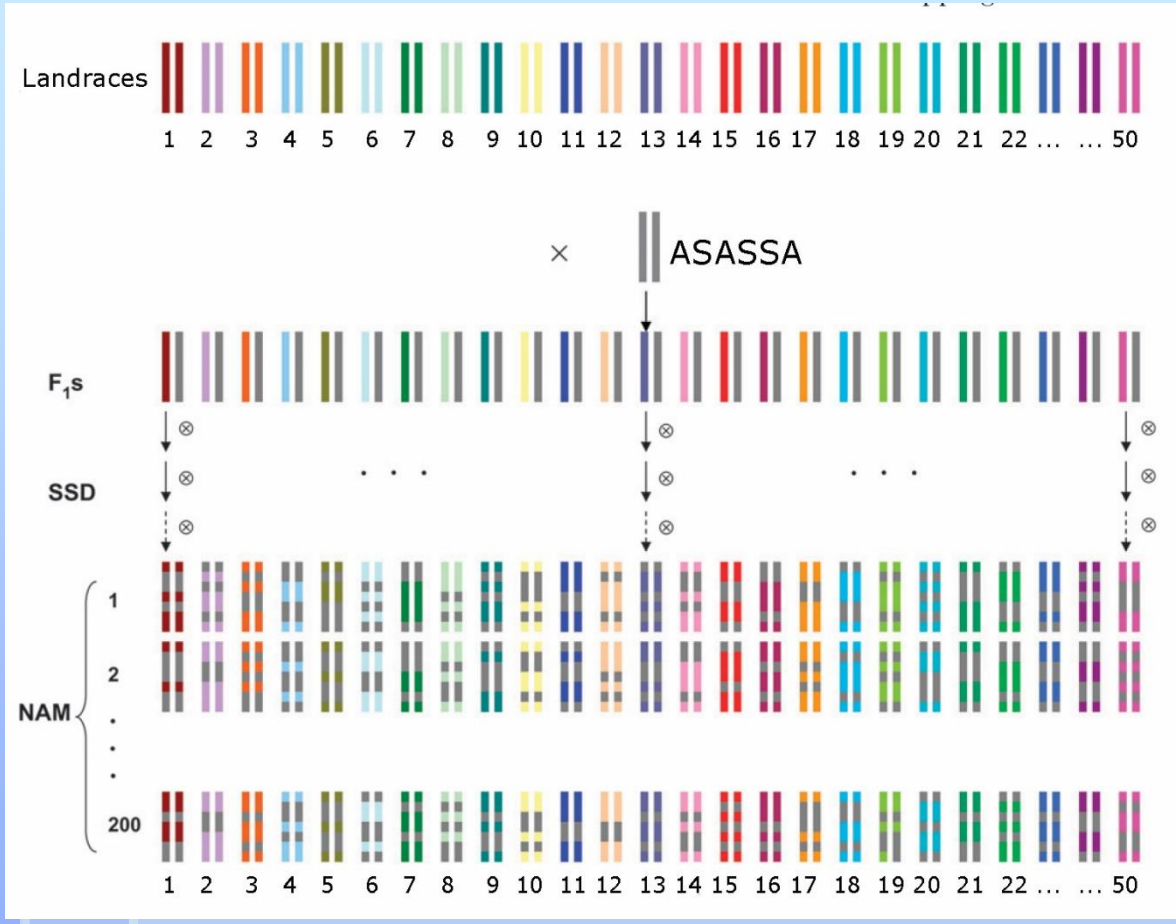
1. **The amount of genetic/phenotypic DIVERSITY present in the population**

   More diversity means more alleles into play

2. **The amount of RECOMBINATIONS (definition by which QTL are detected)**

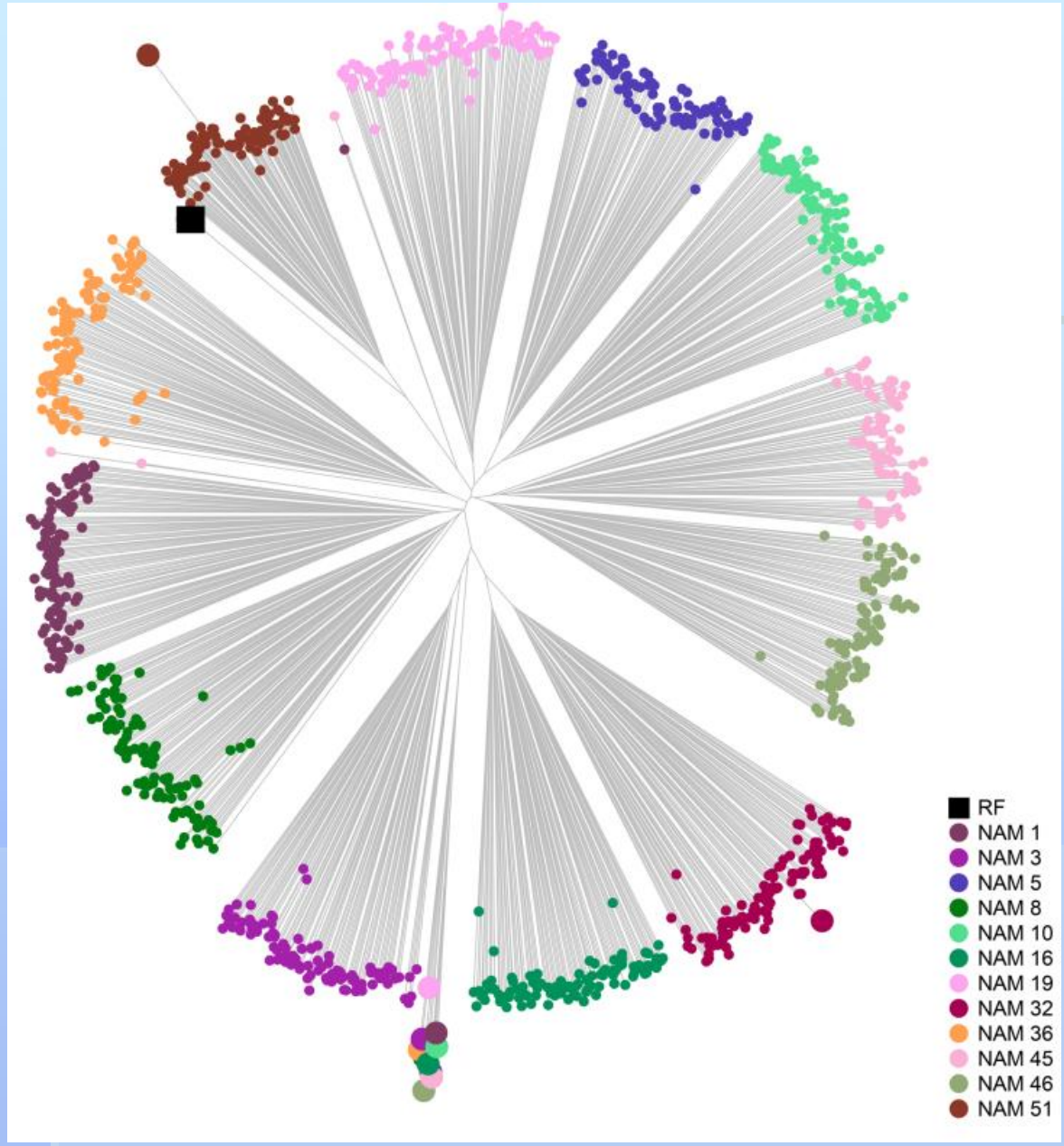
Bi-parental RILs are poor in both
1) Variability
2) N°of informative recombinations

Nested-Association Mapping – NAM



Production of several bi-parerental RIL families sharing a common parent

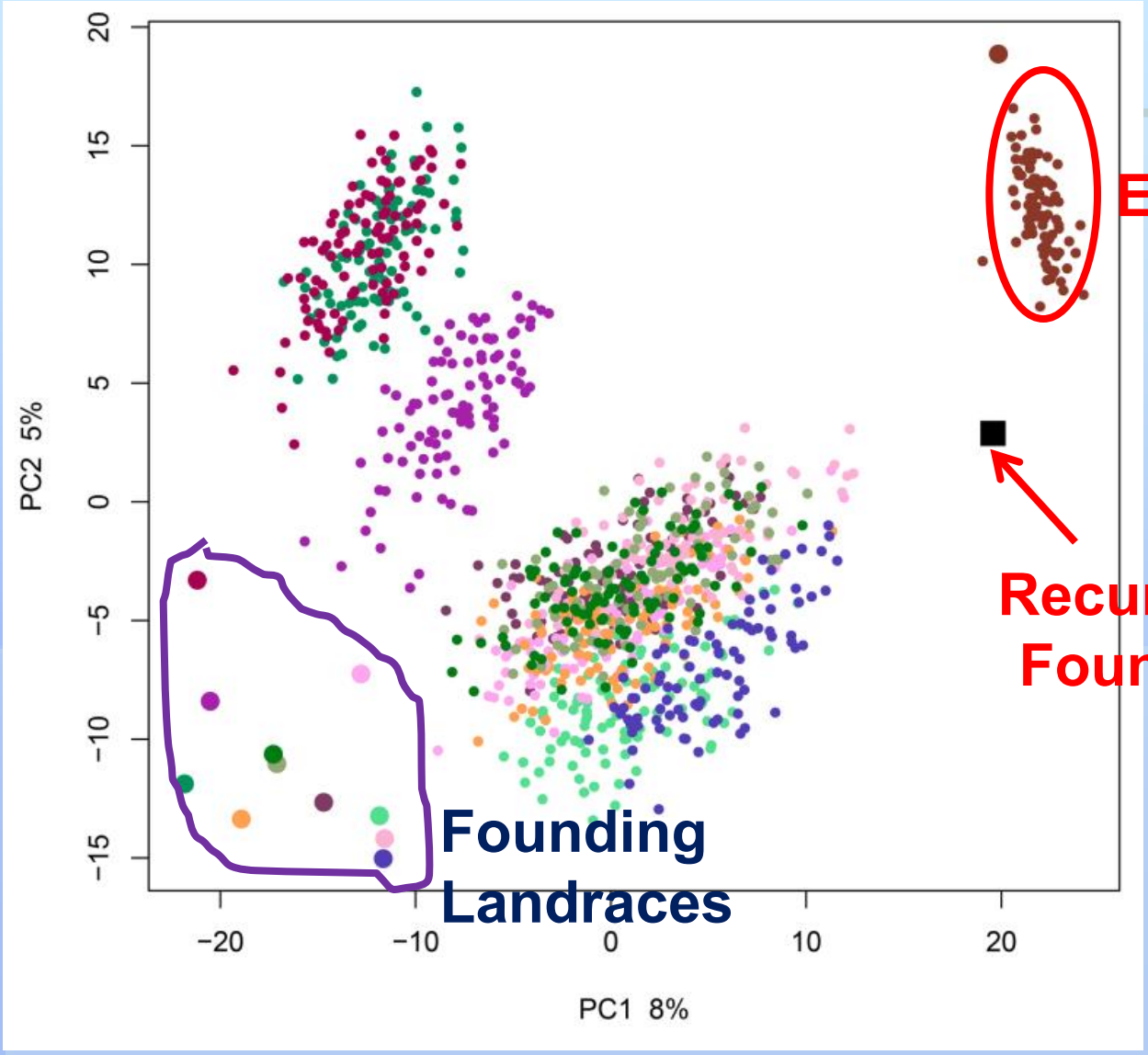
> 6,000 new lines



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

Neighbor Joining Tree 12 monophyletic clades

Filling the Gap between the Ethiopian Landraces and Elite Germplasm



NAM POPULATION CHARACTERISTICS

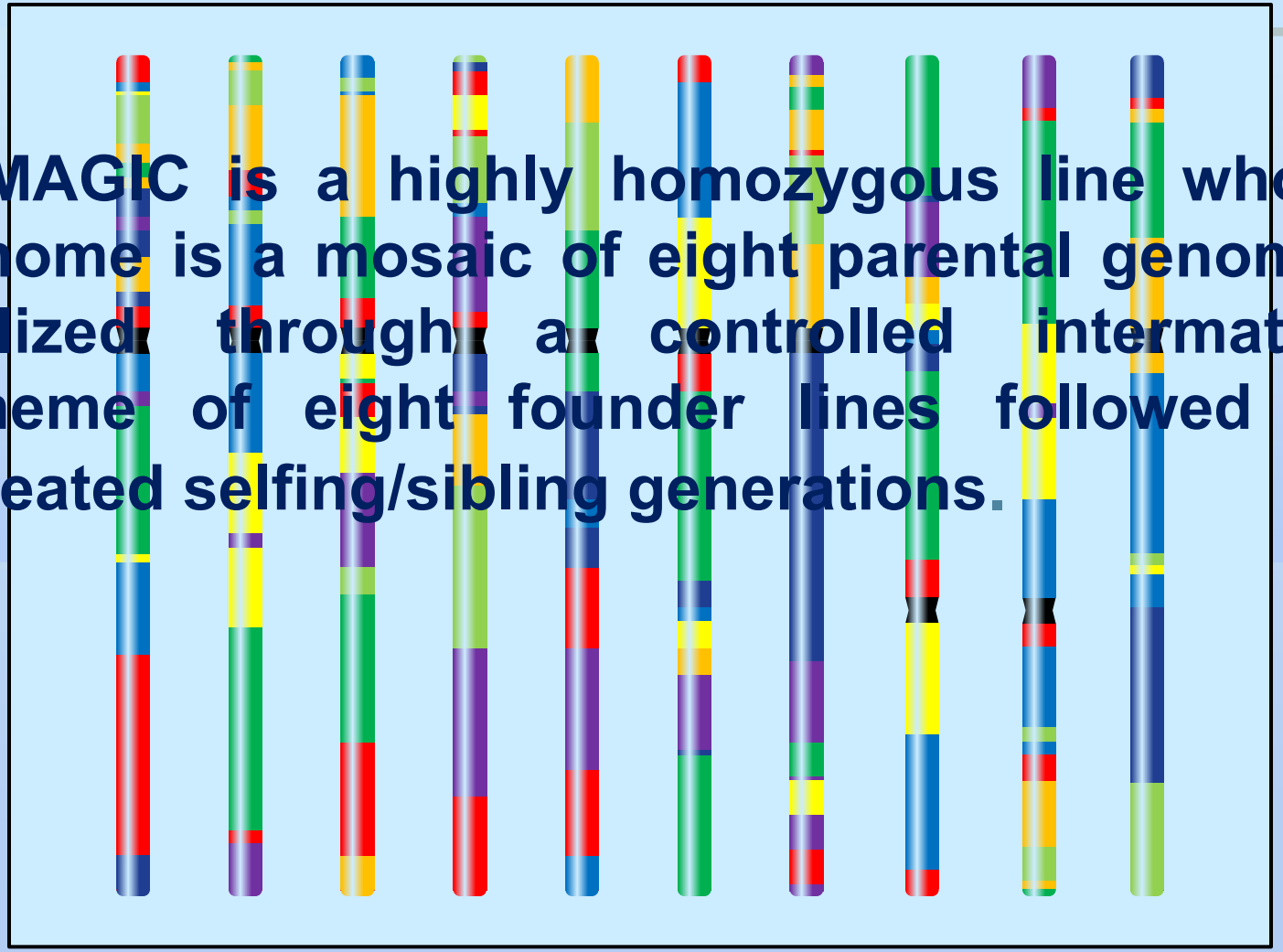
Advantages

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

Disadvantages

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

Multi-parent Advanced Generation of Intercross – MAGIC –



A MAGIC is a highly homozygous line whose genome is a mosaic of eight parental genomes realized through a controlled intermating scheme of eight founder lines followed by repeated selfing/sibling generations.

Original Idea: The Collaborative Cross in Mouse

The Collaborative Cross, a community resource for the genetic analysis of complex traits

The Complex Trait Consortium*

The goal of the Complex Trait Consortium is to promote the development of resources that can be used to understand, treat and ultimately prevent pervasive human diseases. Existing and proposed mouse resources that are optimized to study the actions of isolated genetic loci on a fixed background are less effective for studying intact polygenic networks and interactions among genes, environments, pathogens and other factors. The Collaborative Cross will provide a common reference panel specifically designed for the integrative analysis of complex systems and will change the way we approach human health and disease.

The most common and pervasive human health problems are caused by diseases with complex etiologies. Humans differ greatly in their genetic vulnerability to these common diseases. Mechanisms that underlie disease susceptibility and progression are, with few exceptions, influenced by numerous genetic, developmental and environmental factors. Progress toward treatment and prevention will require new approaches to the genetic analysis of complex systems[1]. We believe that

The Complex Trait Consortium will provide the research community with unique resources needed to discover and dissect the exact contributions of genetic, environmental and developmental components to the etiology of common, complex human diseases. Here we propose a new community resource called the Collaborative Cross that is designed to support the development of integrated models of complex traits. A number of mouse resources for the genetic dissection of disease-

model organism that are based on isolated and transient crosses. By providing a large, common set of genetically defined mice, the Collaborative Cross will become a focal point for cumulative and integrated data collection (Fig. 1), giving rise to a new view of the mammalian organism as a whole and interconnected system.

RI mouse strains are an ideal genetic resource. They can produce unlimited numbers of genetically identical mice that can be

The MAGIC POPULATION
Mixing and Inbreeding

The MAGIC design consists of two main stages:

1. MIXING STAGE: progenitor lines are intercrossed to produce a foundation population whose genomes each contain some genetic material from every founder line

2. INBREEDING STAGE: randomly chosen individuals of the foundation population are inbred (by selfing or sibling), each producing a distinct 8-way recombinant inbred line

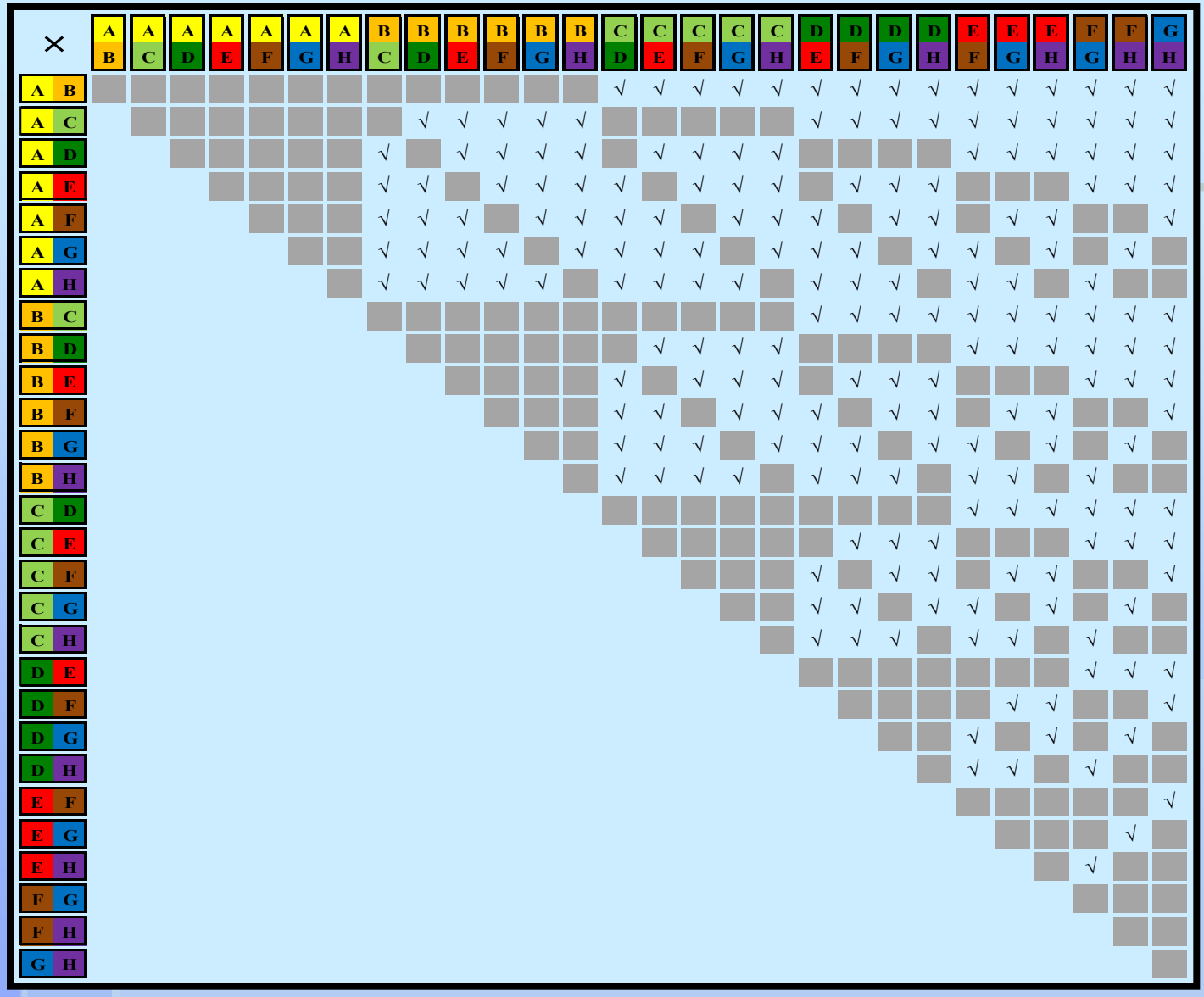
MAGIC Maize: Eight inbred lines selected to develop the population

Inbred line	Group	Subgroup
A632	*SS*	B14A
B73	*SS*	B73
B96	*TS*	Suwan
F7	*Mixed*	-
H99	*NSS*	NSS-mixed
HP301	*Popcorn*	-
Mo17	*NSS*	CO109:Mo17
W153R	*NSS*	NSS-mixed

×	A	B	C	D	E	F	G	H
A		AB	AC	AD	AE	AF	AG	AH
B			BC	BD	BE	BF	BG	BH
C				CD	CE	CF	CG	CH
D					DE	DF	DG	DH
E						EF	EG	EH
F							FG	FH
G								GH
H								

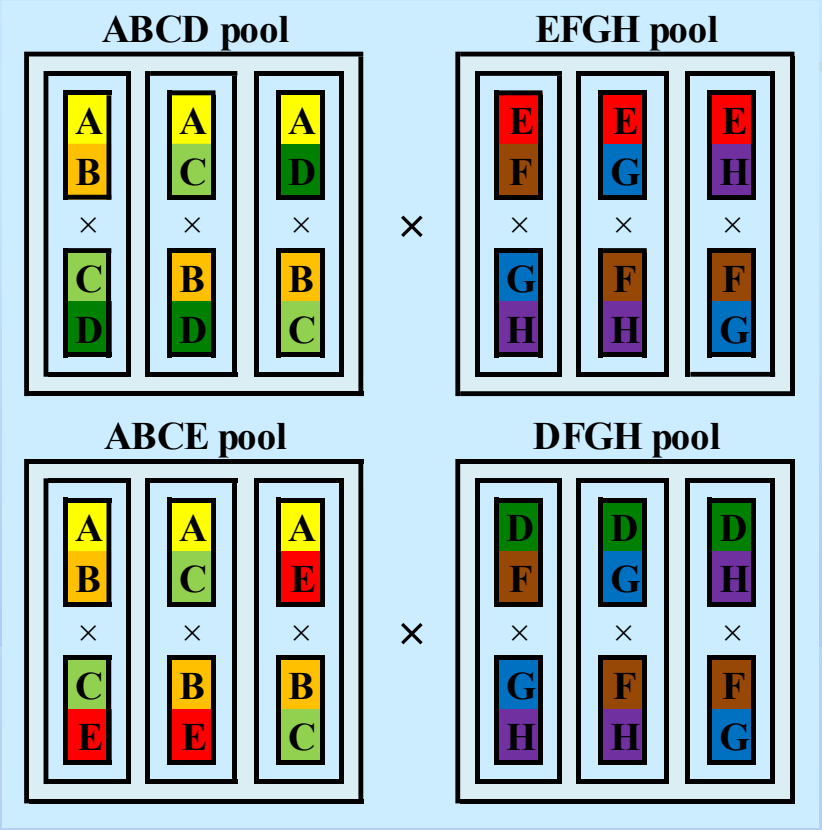
Source: Liu et al. 2003

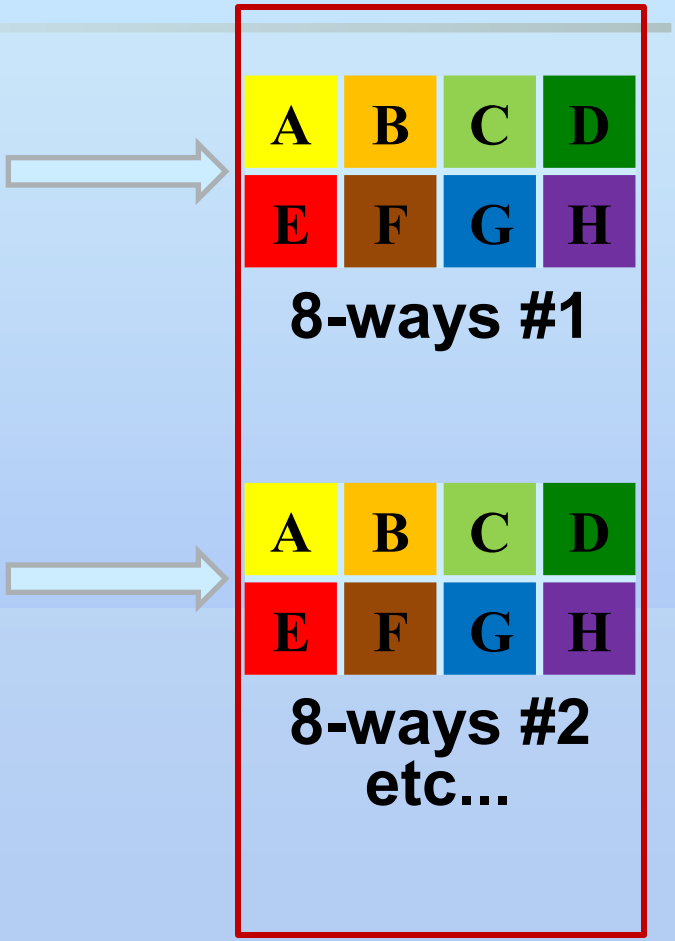
crossed in an incomplete diallel scheme = 28 F1



4 ways are produced crossing hybrid having no parents in common

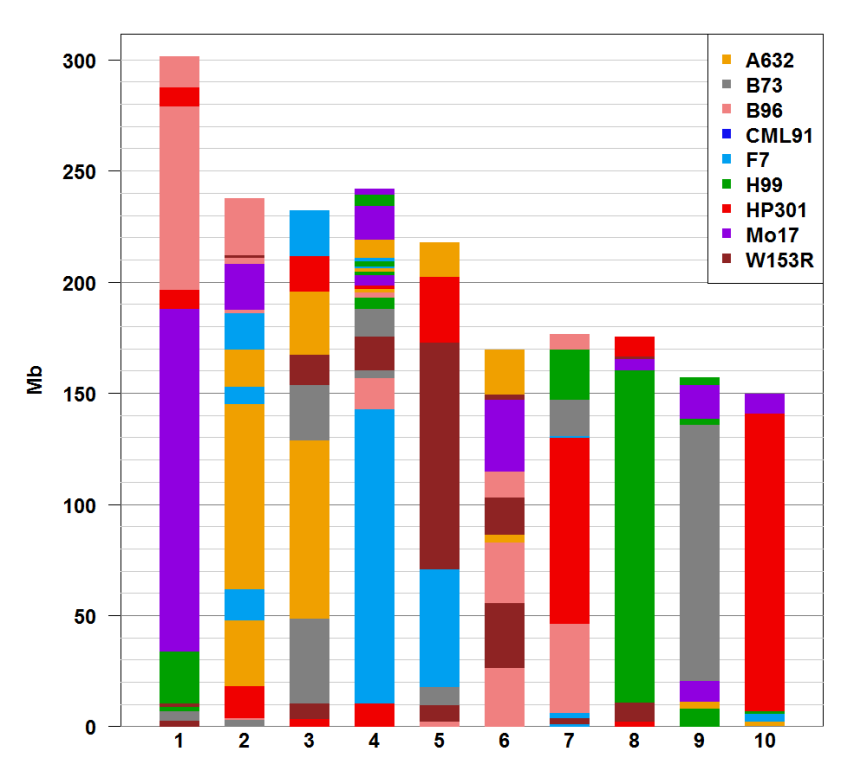
Producing 8-Ways Hybrids

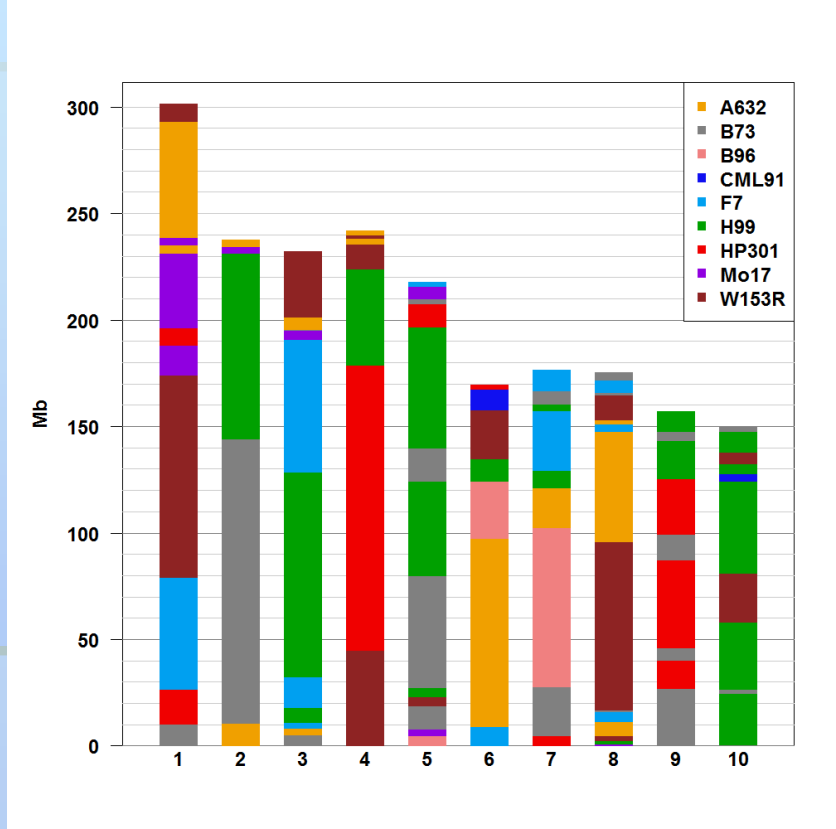




RILs GENOME RECONSTRUCTION

A Hidden Markov Model applied identify the recombination blocks and assigning them to each line





High number of recombination blocks
***i.e.* higher precision mapping**

Additional Layers of Information to Produce a Powerful Platform

Sequencing

Michele Morgante IGA

Genotyping

Matteo Dell'Acqua

Phenotypes

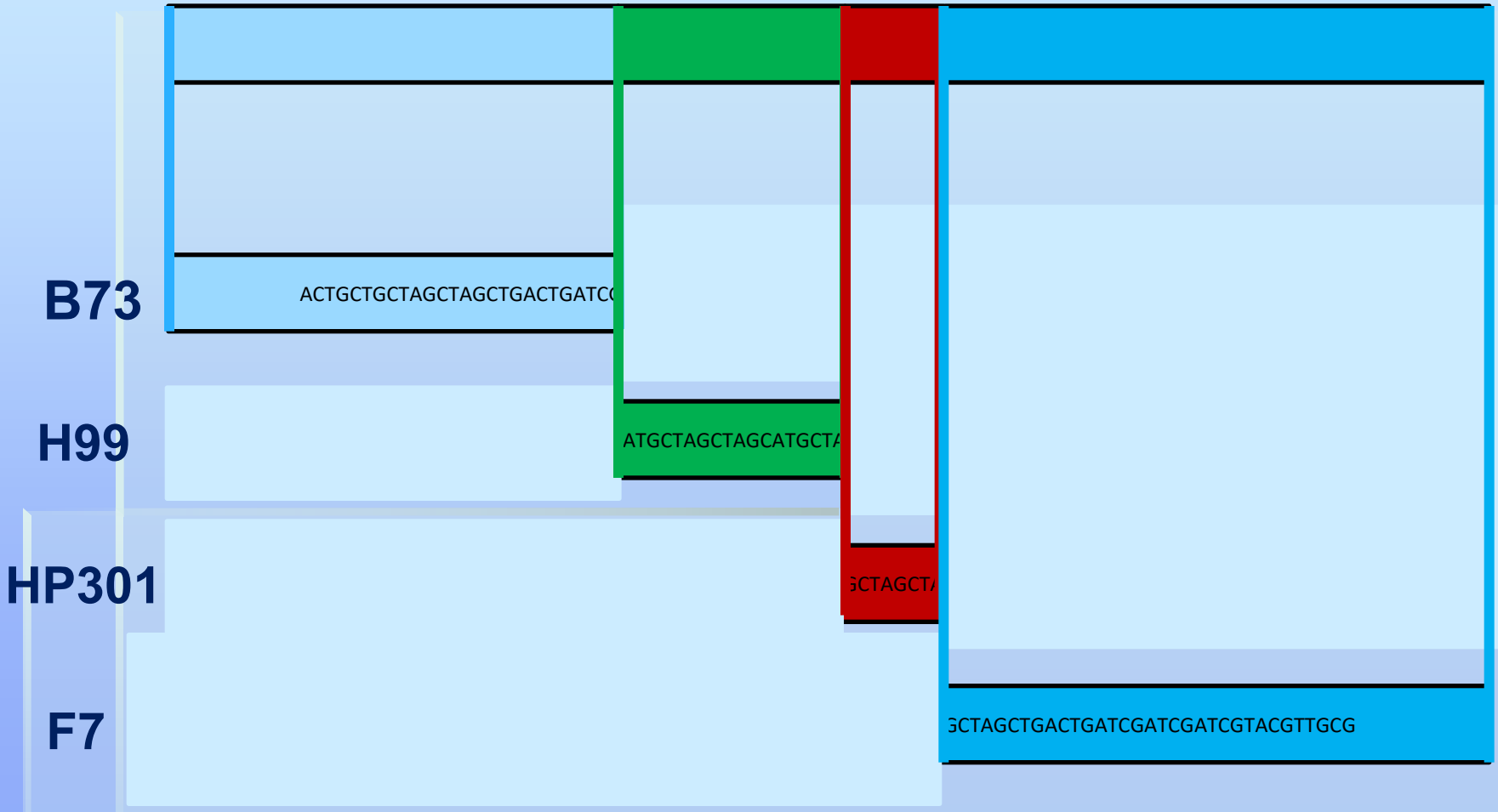
Elisabetta Frascaroli UniBo and many others

Transcriptomics data

Dirk Inzé VIB Gent

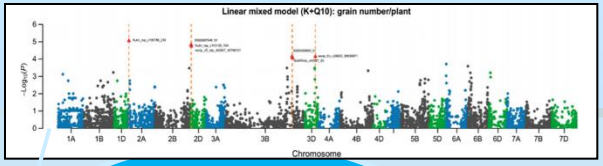
Usefulness of Sequencing the Parents

Reconstructing the genomes – we can use this information to impute sequences on RILs

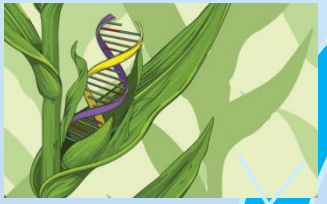


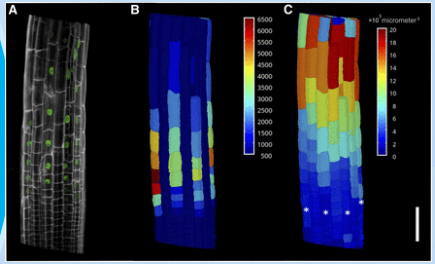
Combining all the pieces in a comprehensive scheme

Genomics
- In field sequencing



Quantitative genetics
- Isolation of genes responsible for adaptation
- Genomic selection

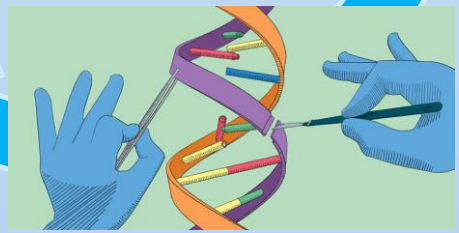




A new era in plant breeding

Molecular Biology
- Systems biology
- Synyetic biology

Precision trait analyses
- Advanced *imaging* in field
- AI



Genome Editing
- Variant substitution