Linkage maps

One of the main uses of DNA markers in agricultural research has been in the construction of **linkage maps** for diverse crop species. Linkage maps have been utilised for identifying chromosomal regions that contain genes controlling simple traits (controlled by a single gene) and quantitative traits using QTL analysis.

What are linkage maps?

A linkage map may be thought of as a 'road map' of the chromosomes derived from two different parents (Paterson, 1996a). Linkage maps indicate the position and relative genetic distances between markers along chromosomes, which is analogous to signs or landmarks along a highway. The most important use for linkage maps is to identify chromosomal locations containing genes and QTLs associated with traits of interest; such maps may then be referred to as 'QTL' or 'genetic' maps.

Genetic mapping is based on the principle that genes and markers segregate via chromosome recombination (called crossing-over) during meiosis (i.e. sexual reproduction), thus allowing their analysis in the progeny (Paterson, 1996a).

Euphytica (2005) 142: 169–196
DOI: 10.1007/s10681-005-1681-5

An introduction to markers, ~~quantitative trait loci (QTL) mapping~~ and marker-assisted selection for crop improvement: The basic concepts

B.C.Y. Collard[1,4,*], M.Z.Z. Jahufer[2], J.B. Brouwer[3] & E.C.K. Pang[1]

A typical output of a linkage map.

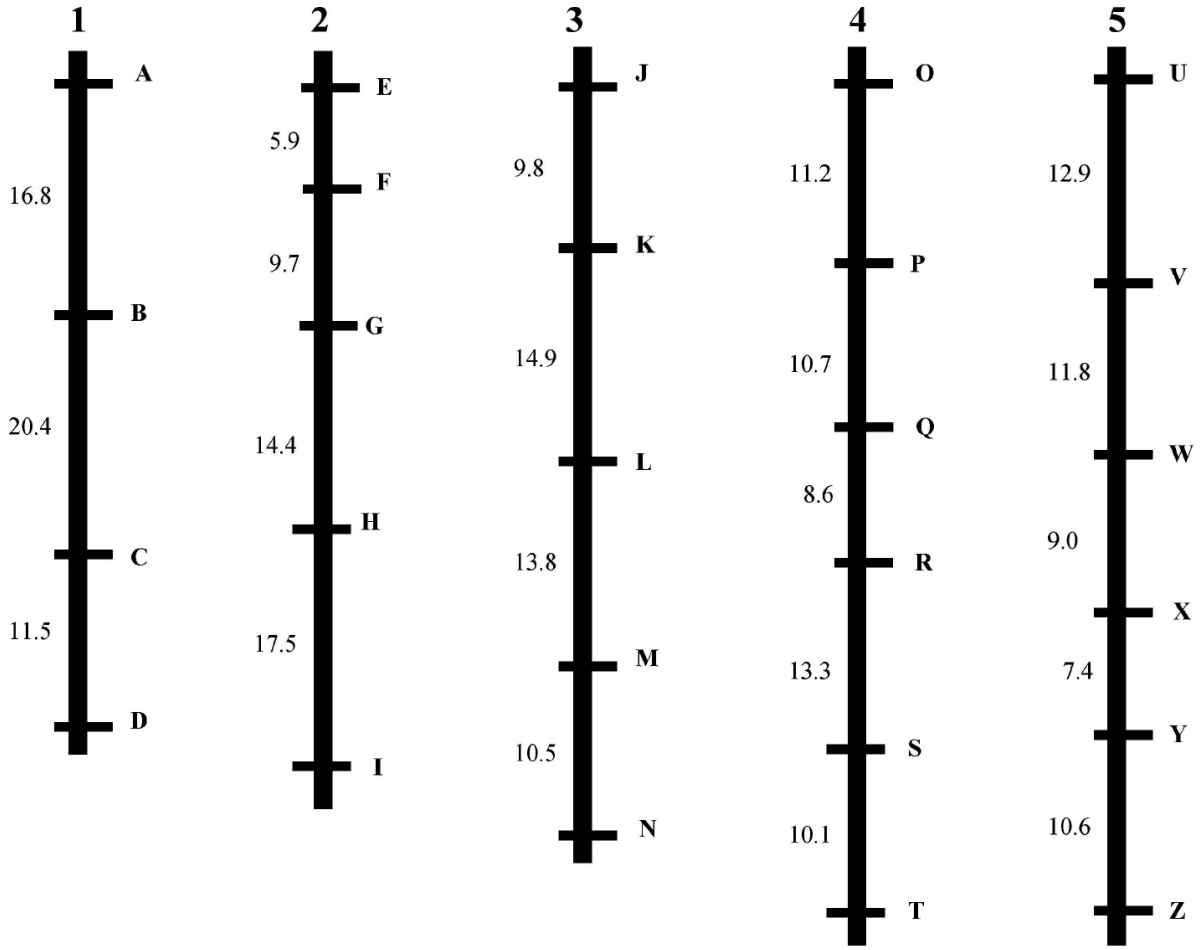


Figure 7. Hypothetical 'framework' linkage map of five chromosomes (represented by linkage groups) and 26 markers. Ideally, a framework map should consist of evenly spaced markers for subsequent , analysis.

Linked markers are grouped together into 'linkage groups,' which represent chromosomal segments or entire chromosomes

Genes or markers that are close together or tightly-linked will be transmitted together from parent to progeny more frequently than genes or markers that are located further apart (Figure 3). In a segregating population, there is a mixture of parental and recombinant genotypes.

The frequency of recombinant genotypes can be used to calculate recombination fractions, which may by used to infer the genetic distance between markers. By analysing the segregation of markers, the relative order and distances between markers can be determined–the lower the frequency of recombination between two markers the closer they are situated on a chromosome (conversely, the higher the frequency of recombination between two markers, the further away they are situated on a chromosome).

Mapping functions are used to convert recombination fractions into map units called centi-Morgans (cM). 1 cM corresponds to a recombination frequency of 1%.

Markers that have a recombination frequency of 50% are described as ‘unlinked’ and assumed to be located far apart on the same chromosome or on different chromosomes.

Linkage maps are constructed from the analysis of many segregating markers.

The three main steps of linkage map construction are:
(1) production of a mapping population;
(2) identification of polymorphism and
(3) linkage analysis of markers.

MAPPING POPULATIONS

Mapping populations

The construction of a linkage map requires a segregating plant population (i.e. a population derived from sexual reproduction).

The parents selected for the mapping population will differ for one or more traits of interest. Population sizes used in preliminary genetic mapping studies generally range from 50 to 250 individuals.

However larger populations are required for high-resolution mapping.

Generally in self-pollinating species, mapping populations originate from parents that are both highly homozygous (inbred). In cross pollinating species, the situation is more complicated since most of these species do not tolerate inbreeding. Many cross pollinating plant species are also polyploid (contain several sets of chromosome pairs). Mapping populations used for mapping cross pollinating species may be derived from a cross between a heterozygous parent and a haploid or homozygous parent (Wu et al., 1992). For example, in both the cross pollinating species white clover (*Trifolium repens* L.) and ryegrass (*Lolium perenne* L.), F1 generation mapping populations were successfully developed by pair crossing heterozygous parental plants that were distinctly different for important traits associated with plant persistence and seed yield.

Several different populations may be utilized for mapping within a given plant species, with each population type possessing advantages and disadvantages

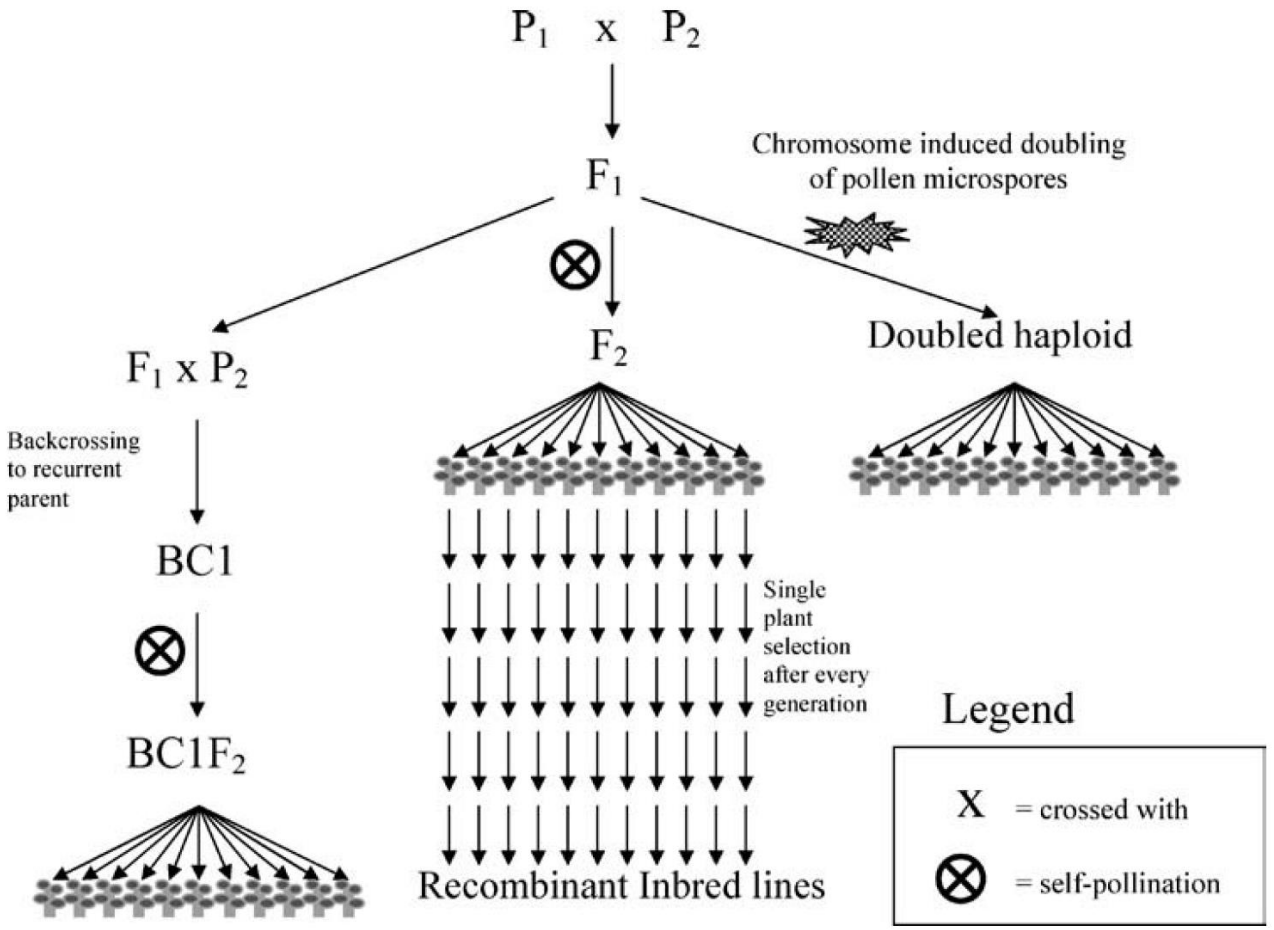


Figure 4. Diagram of main types of mapping populations for self-pollinating species.

F2 populations, derived from F1 hybrids, and backcross (BC) populations, derived by crossing the F1 hybrid to one of the parents, are the simplest types of mapping populations developed for self pollinating species. Their main advantages are that they are easy to construct and require only a short time to produce.

Recombinant inbred (RI) lines, which consist of a series of homozygous lines, each containing a unique combination of chromosomal segments from the original parents, can be generated by i**nbreeding from individual F2 plants** .
The length of time needed for producing RI populations is the major disadvantage, because usually six to eight generations are required.

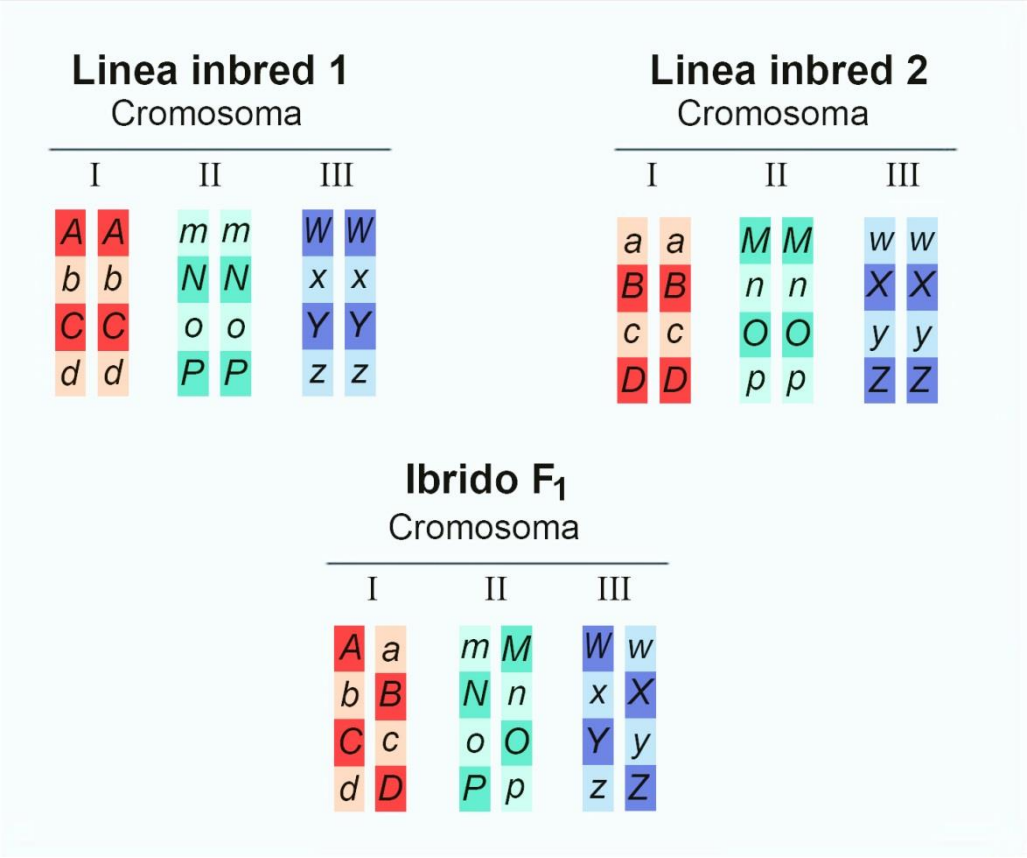
Doubled haploid (DH) populations may be produced by regenerating plants by the induction of chromosome doubling from pollen grains, however the production of DH populations is only possible in species that are amenable to tissue culture (e.g. cereal species such as rice, barley and wheat).

The major advantages of RI and DH populations are that they produce homozygous or 'true-breeding' lines that can be multiplied and reproduced without genetic change occurring. This allows for the conduct of replicated trials across different locations and years. Thus both RI and DH populations represent 'eternal' resources for genetic mapping.
Furthermore, seed from individual RI or DH lines may be transferred between different laboratories for further linkage analysis and the addition of markers to existing maps, ensuring that all collaborators examine identical materialRI or DH lines may be transferred between different laboratories for further linkage analysis and the addition of markers to existing maps, ensuring that all collaborators examine identical material.

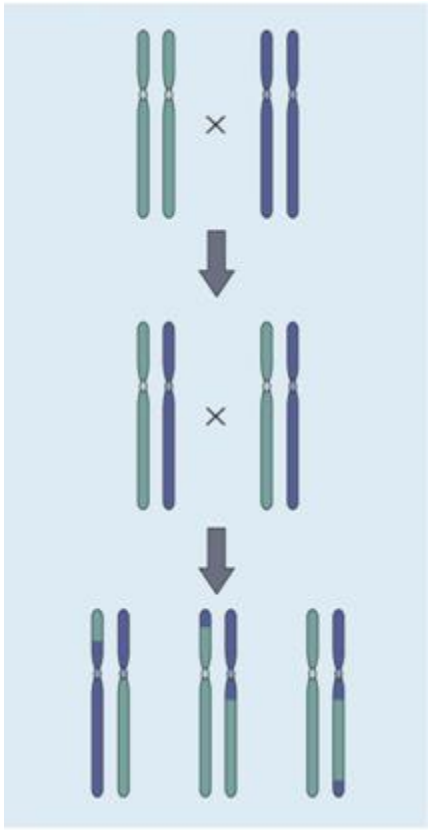
PRODUZIONE DI IBRIDI F1

Possono essere ottenuti sia da piante autogame che allogame

1) Costituzione di linee INBRED (pure) per autoimpollinazione
2) Impollinazione incrociata tra linee inbred selezionate per ottenere ibrido F1



F2 populations



$$P_1(AA) \quad x \quad P_2(aa)$$

$$\downarrow$$

$$F_1(Aa) \quad x \quad F_1(Aa)$$

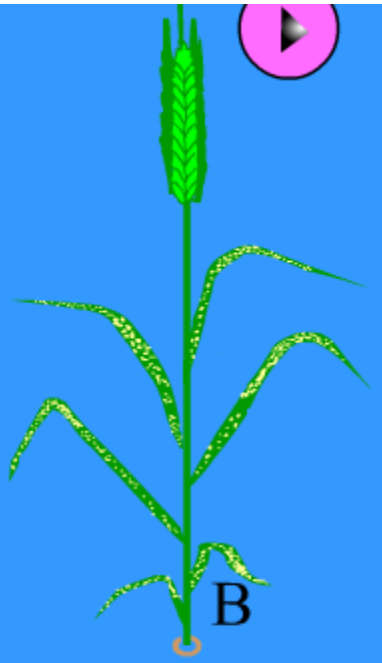
$$\downarrow$$

$$F_2(AA, Aa, aa)$$

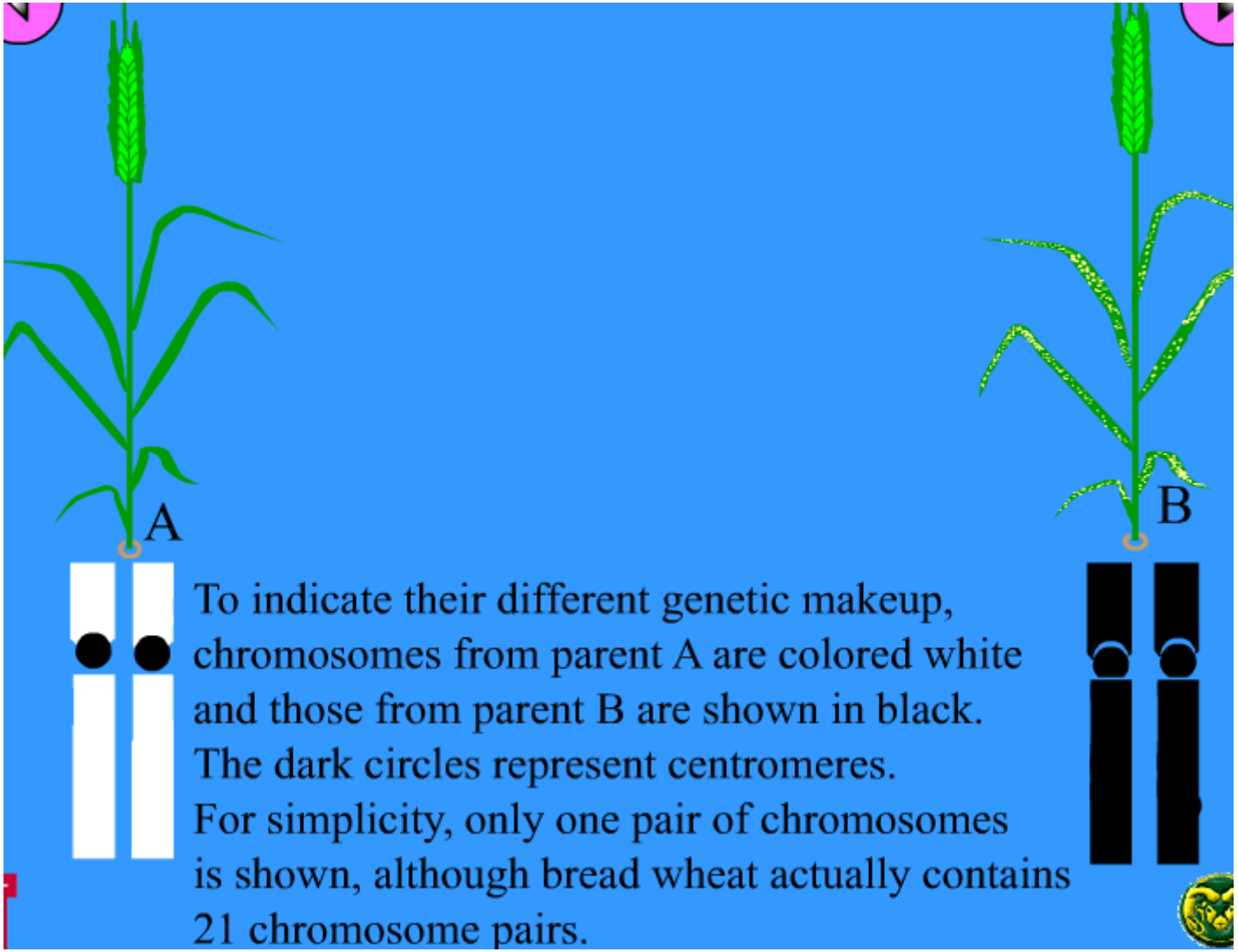
Advantage: Fast and easy to construct

Disadvantage: F3 families are still very heterozygous, so the precision of the estimates can be low (because of the high standard error); can't be replicated

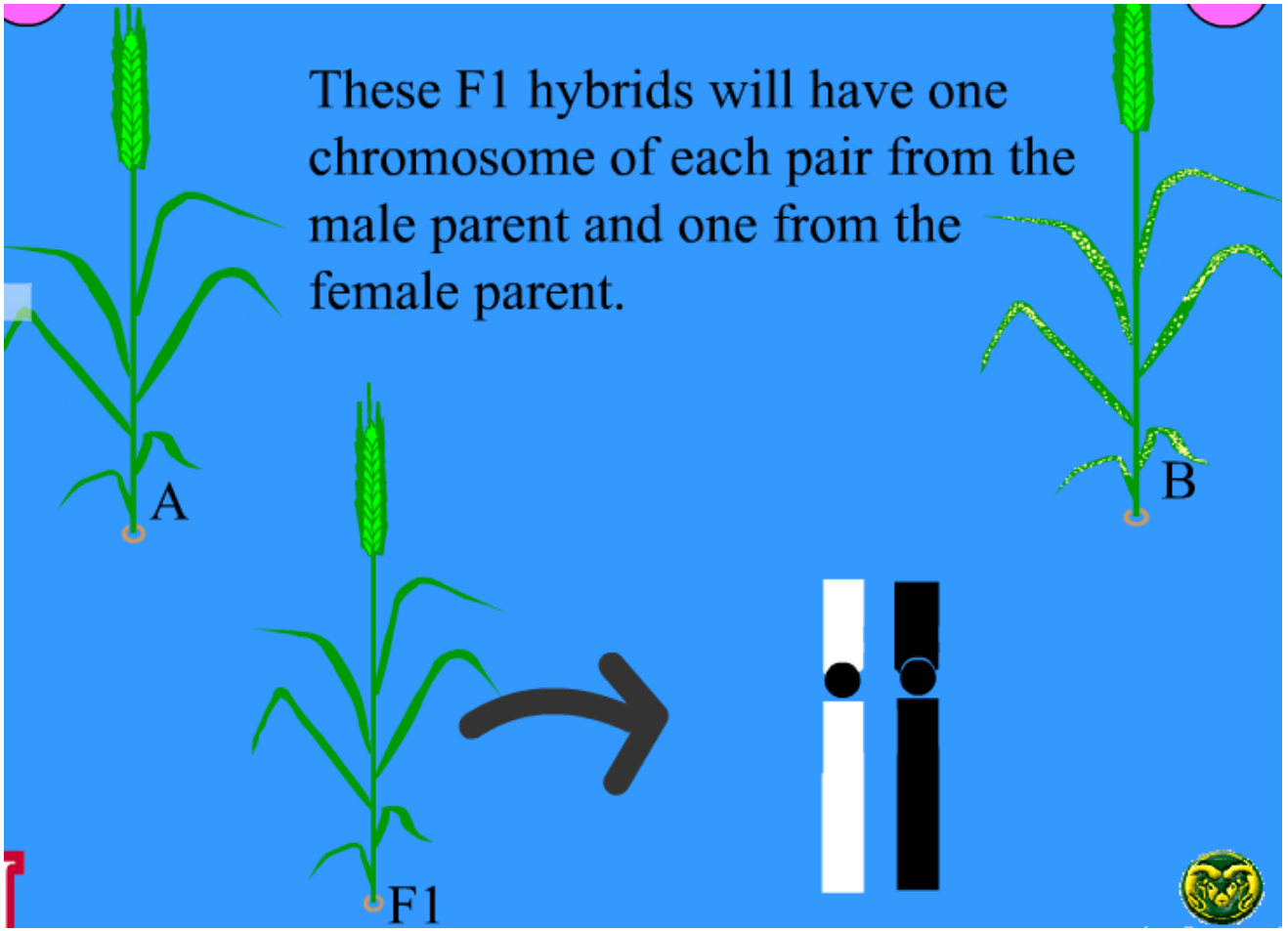
Jampatong *et al.* (2002)



The first step is to cross parent lines A (resistant) and B (susceptible). This involves placing pollen from one parent onto the stigmas of the other parent to produce hybrid seed.

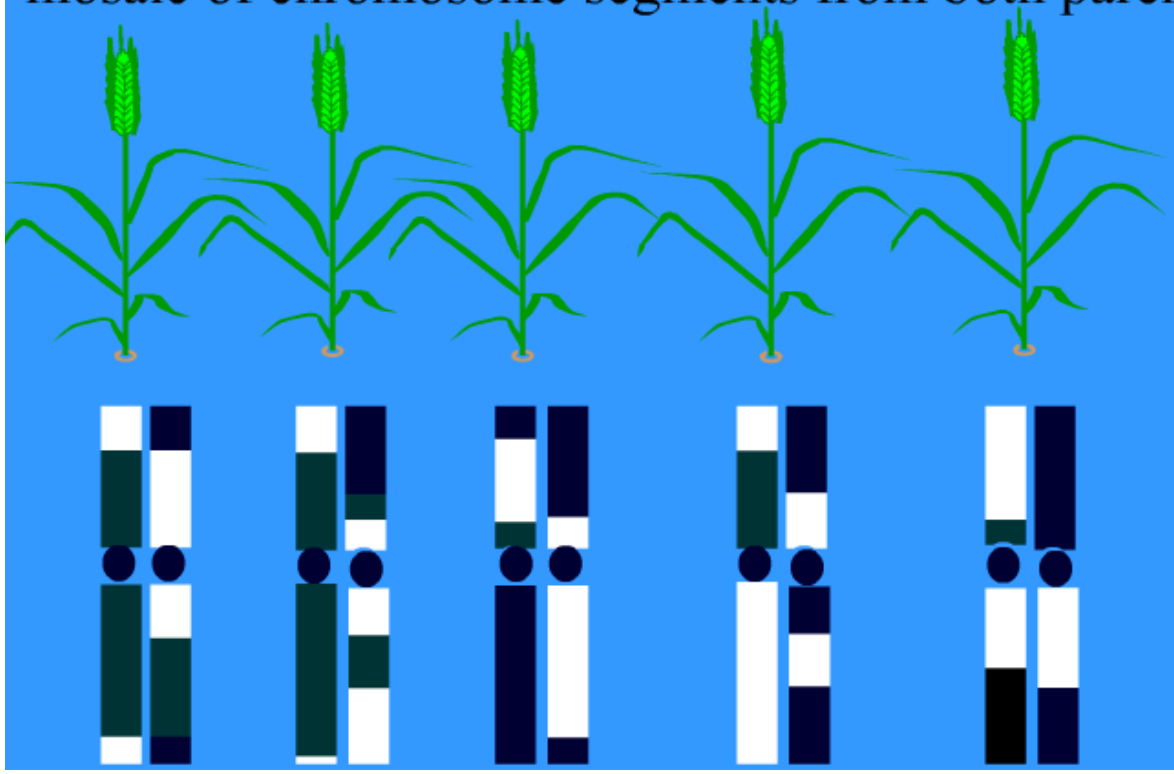

A
B
To indicate their different genetic makeup,
chromosomes from parent A are colored white
and those from parent B are shown in black.
The dark circles represent centromeres.
For simplicity, only one pair of chromosomes
is shown, although bread wheat actually contains
21 chromosome pairs.

F1 hybrid plants are grown from seed resulting from the cross of A×B.



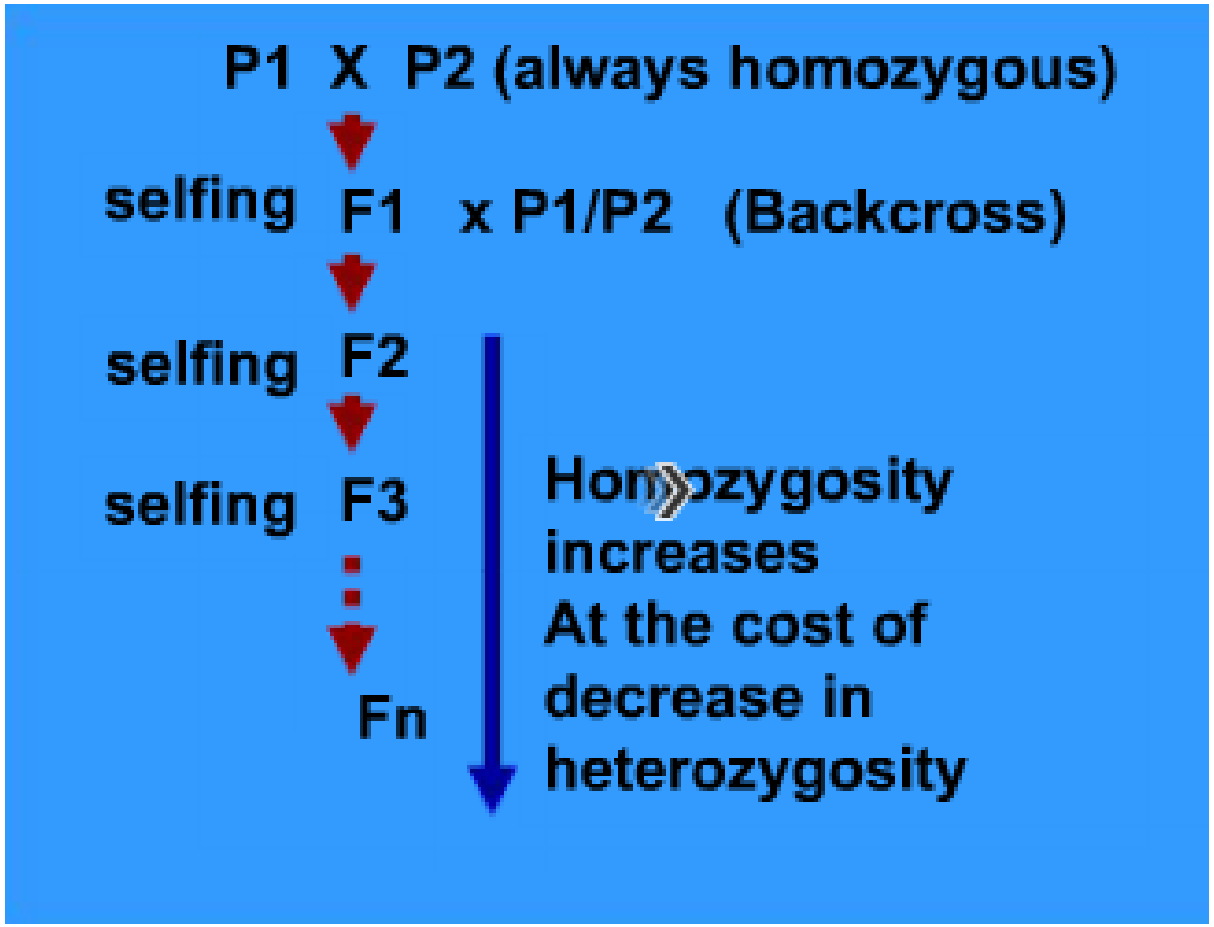
Self-pollinating an F1 plant produces an array of F2 plants.

Due to genetic recombination that occurs in the reproductive process, each F2 plant is a unique mosaic of chromosome segments from both parents.



The F2 population is one type of population used for QTL mapping. Typical population size is 100-300 individual F2 plants, rather than the few plants shown here.

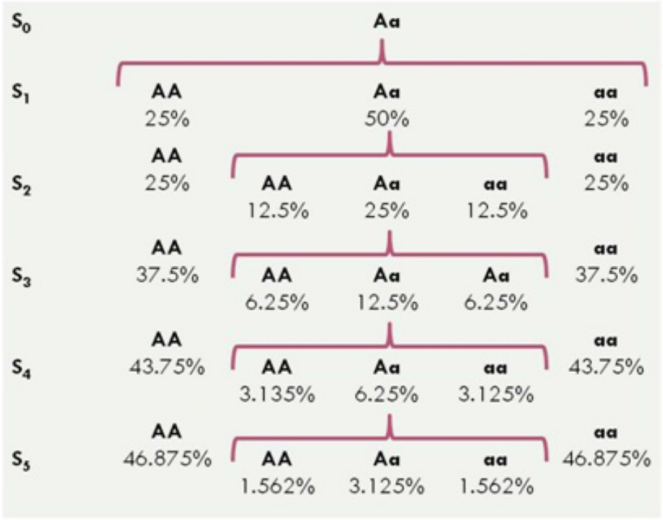
- Consequences of selfing in:
  - Cross-pollinated crops



Quantificare l'incremento dell'omozigosi: popolazioni autogame

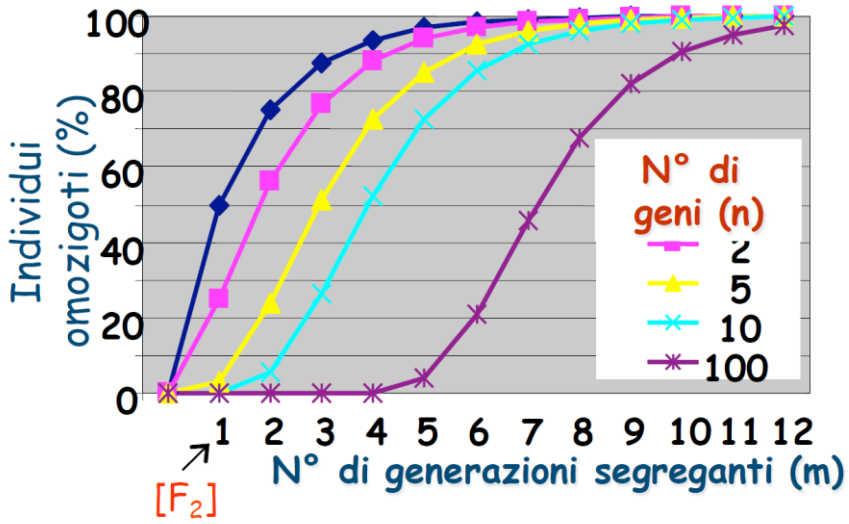
- Assumiamo una popolazione S_0 che non sia il risultato di inbreeding (es. una popolazione di ibridi F_1 derivante da incrocio tra linee pure). Per quanto detto, alla prima generazione di autofecondazione (S_1), $F_{S1}=\frac{1}{2}$. Dopo due generazioni di autofecondazione, $F_{S2}=\frac{1}{2}(1+\frac{1}{2})=\frac{3}{4}$, dopo 3 generazioni $F_{S3}=7/8$, e cosi via, fino a tendere come detto a $F^*=1$.

- Considerato che f(Aa)=2(1-F)pq, si ha il risultato che la quota di eterozigosi si dimezza ad ogni generazione di autofecondazione. Le popolazioni naturali di specie autogame sono un insieme di differenti linee pure

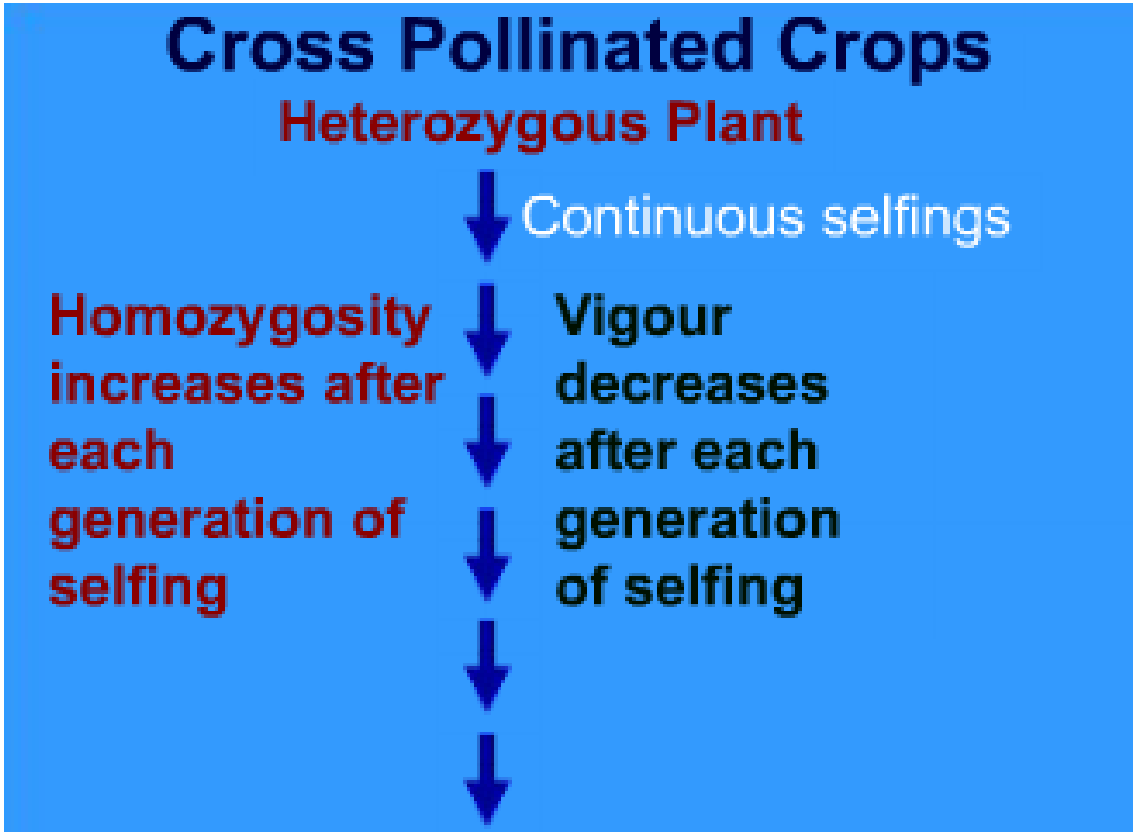


Incremento dell'omozigosi in popolazioni autogame

Considerato un genotipo eterozigote per n coppie alleliche, ed m generazioni di autofecondazione



$$\left(\frac{2^m - 1}{2^m}\right)^n$$


Cross Pollinated Crops
Heterozygous Plant
Continuous selfings
Homozygosity increases after each generation of selfing
Vigour decreases after each generation of selfing

Backcross (BC)

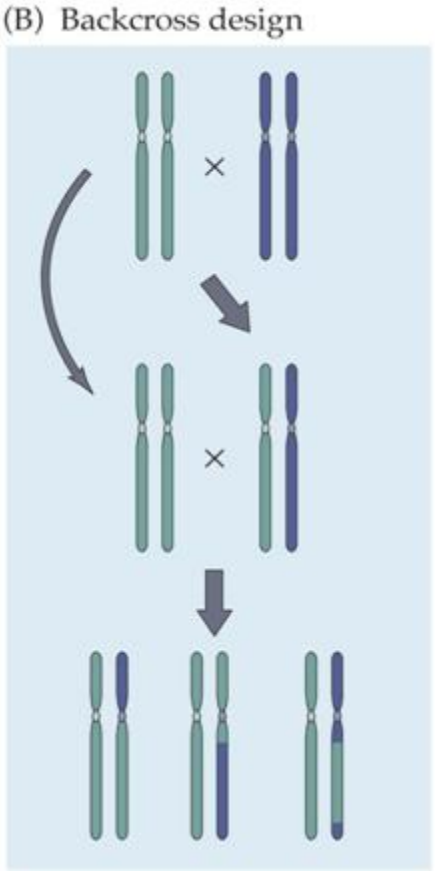
$$P_1(AA) \quad x \quad P_2(aa)$$

$$\downarrow$$

$$P_1(AA) \; x \; F_1(Aa)$$

$$\downarrow$$

$$BC_1$$



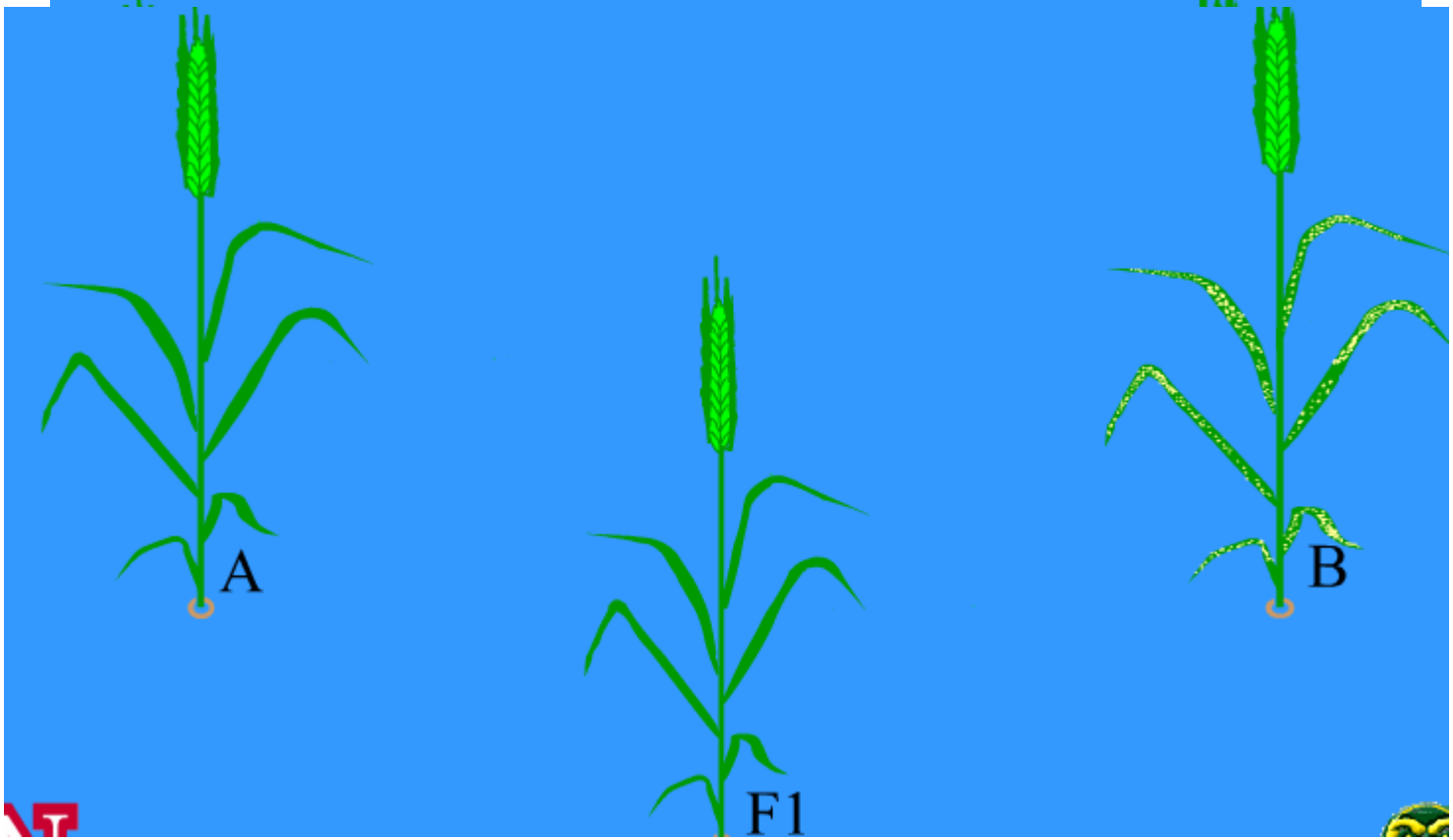
Advantages: It is easier to identify QTL as there are less epistatic and linkage drag effects; especially useful for crosses with wild species.

Disadvantages: Difficult or impossible in species that are highly heterozygous and outcrossing.

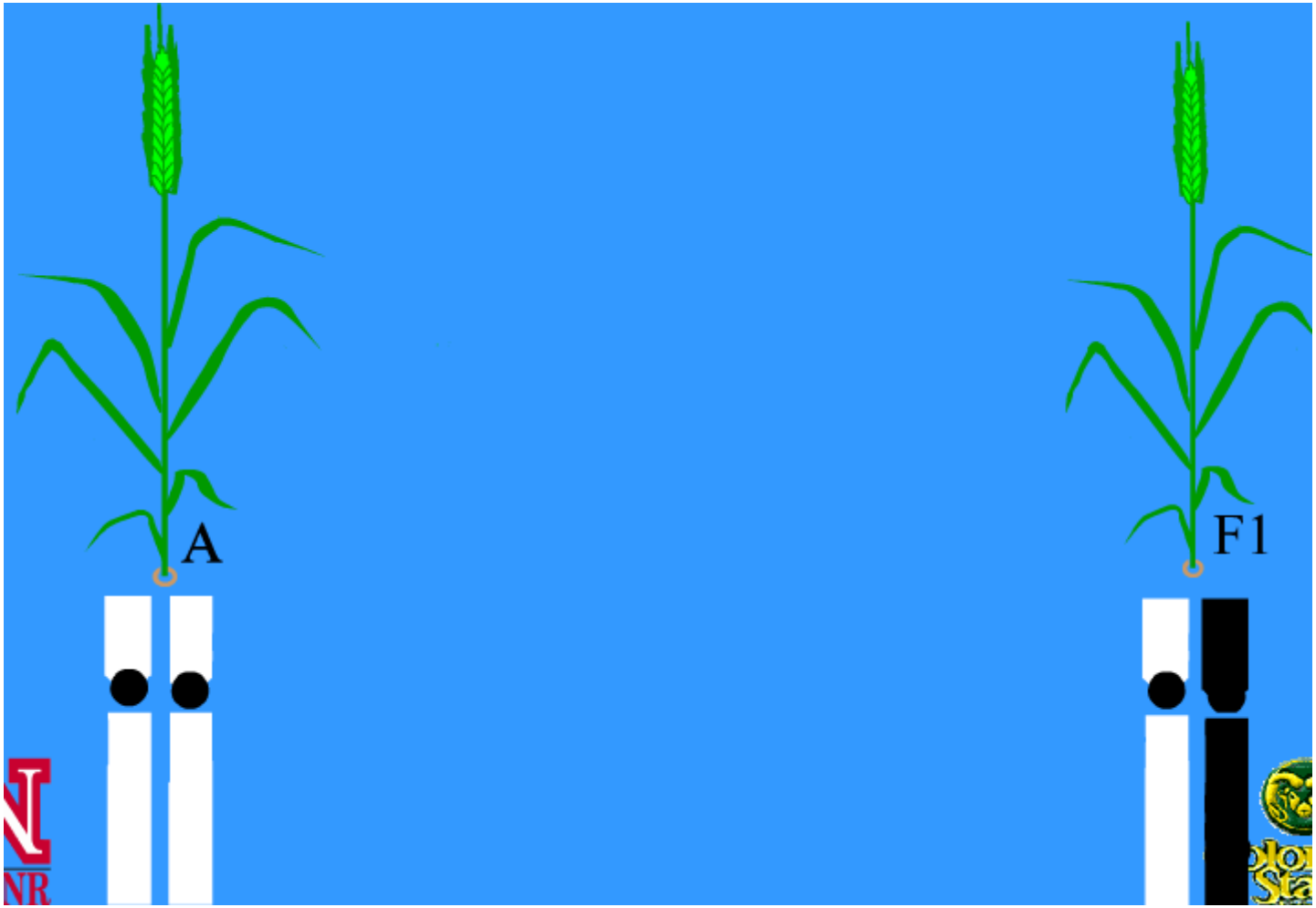
Use: best when inbred lines are available

Huang *et al.(2003)*

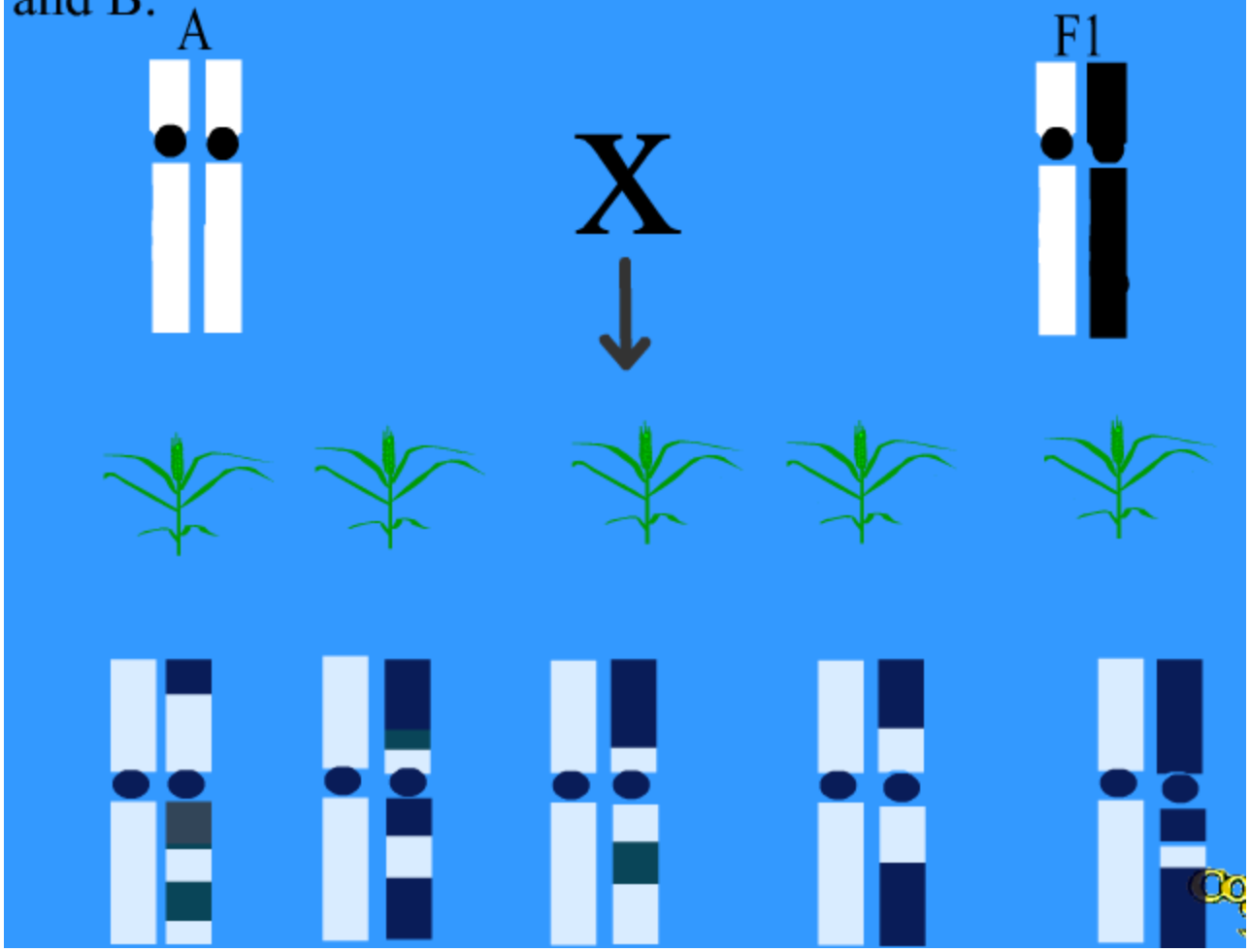
A [redacted] frequently used QTL mapping population is a backcross population. First, an F1 plant is produced by crossing parent lines A and B.



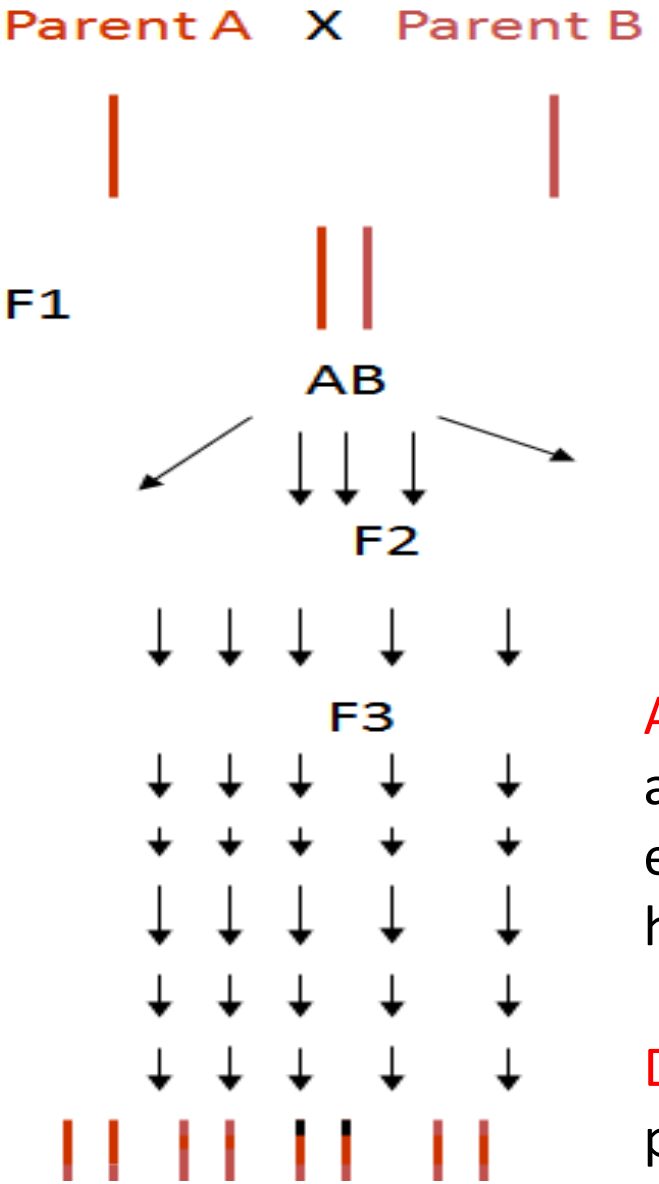
Then the F1 is crossed to one of the parent lines, in this case parent A.



The result is a population of individuals, all of which have one chromosome from parent A and a recombined chromosome containing segments from both parents A and B.



Recombinant inbred (RI) lines



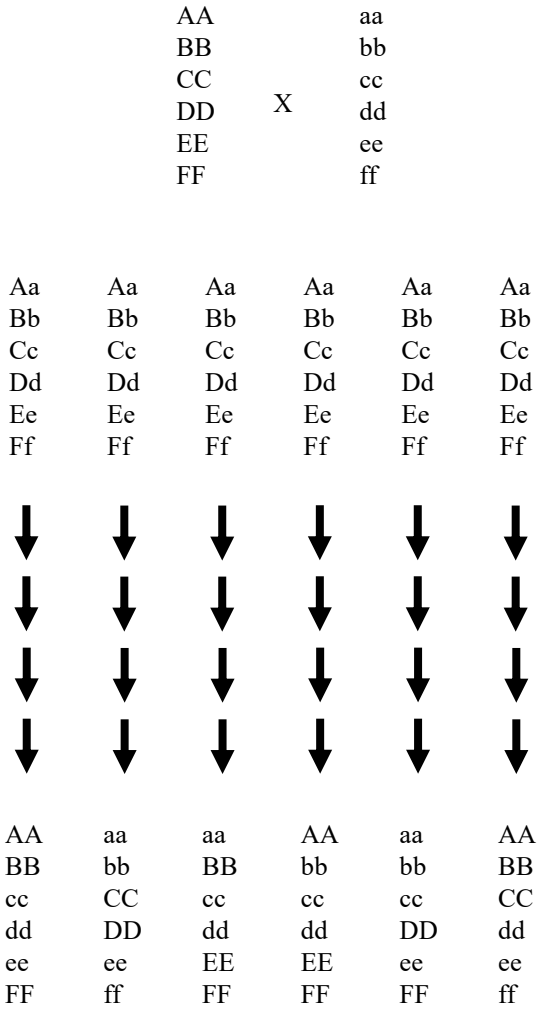
- True breeding or homozygous
- Immortal collection
- Replicate experiments in different environments
- Molecular Marker database can be updated

Advantages: fixed lines so can be replicated across many locations and/or years; can eliminate problem of background heterozygosity

Disadvantages: Can take a long time to produce. (Some species are not amenable).

He P et al.(2001)

Recombinant inbred (RI) lines



Basic method

- Start with inbred parental lines, homozygous at every locus
- Make F1s, heterozygous at every locus
- Inbreed different F1 lines
- These recombinant inbred lines are homozygous at each locus
- Compare strain distribution pattern of markers and traits

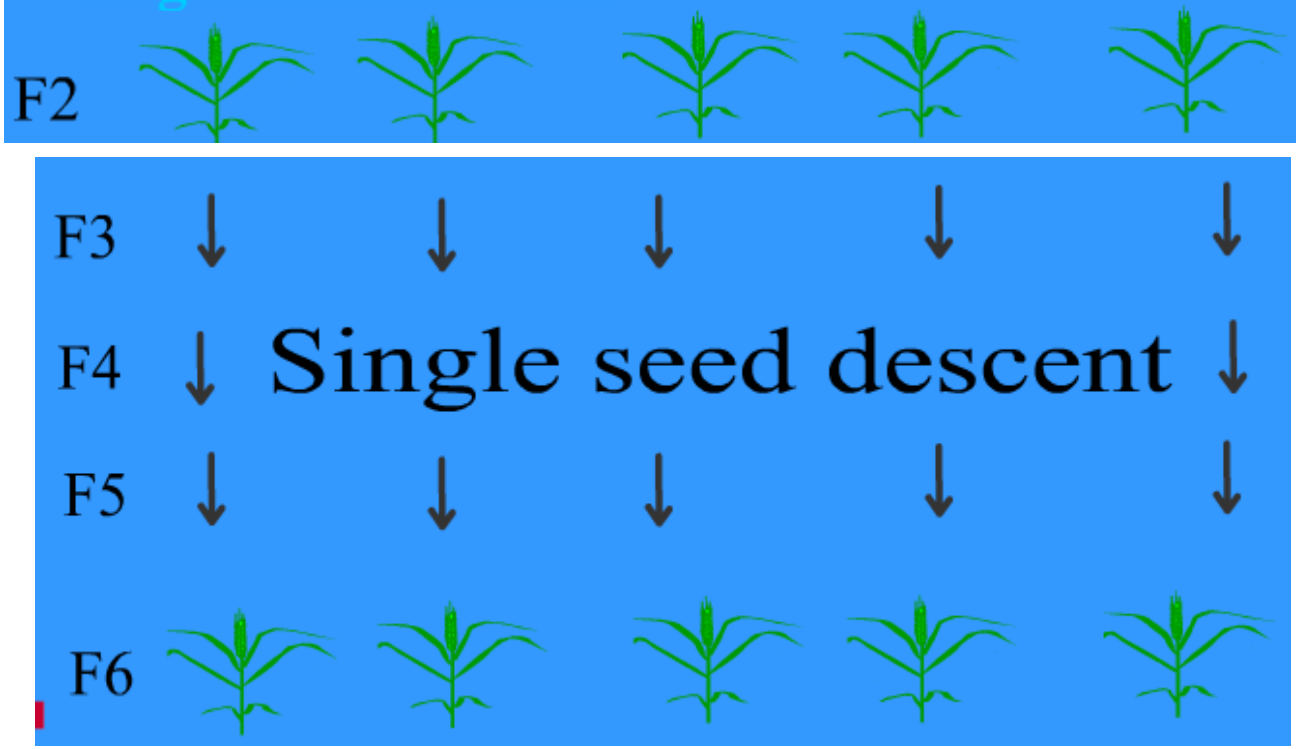
Advantages

•Multiple genetically identical individuals can be scored for each line, decreasing environmental contribution to trait

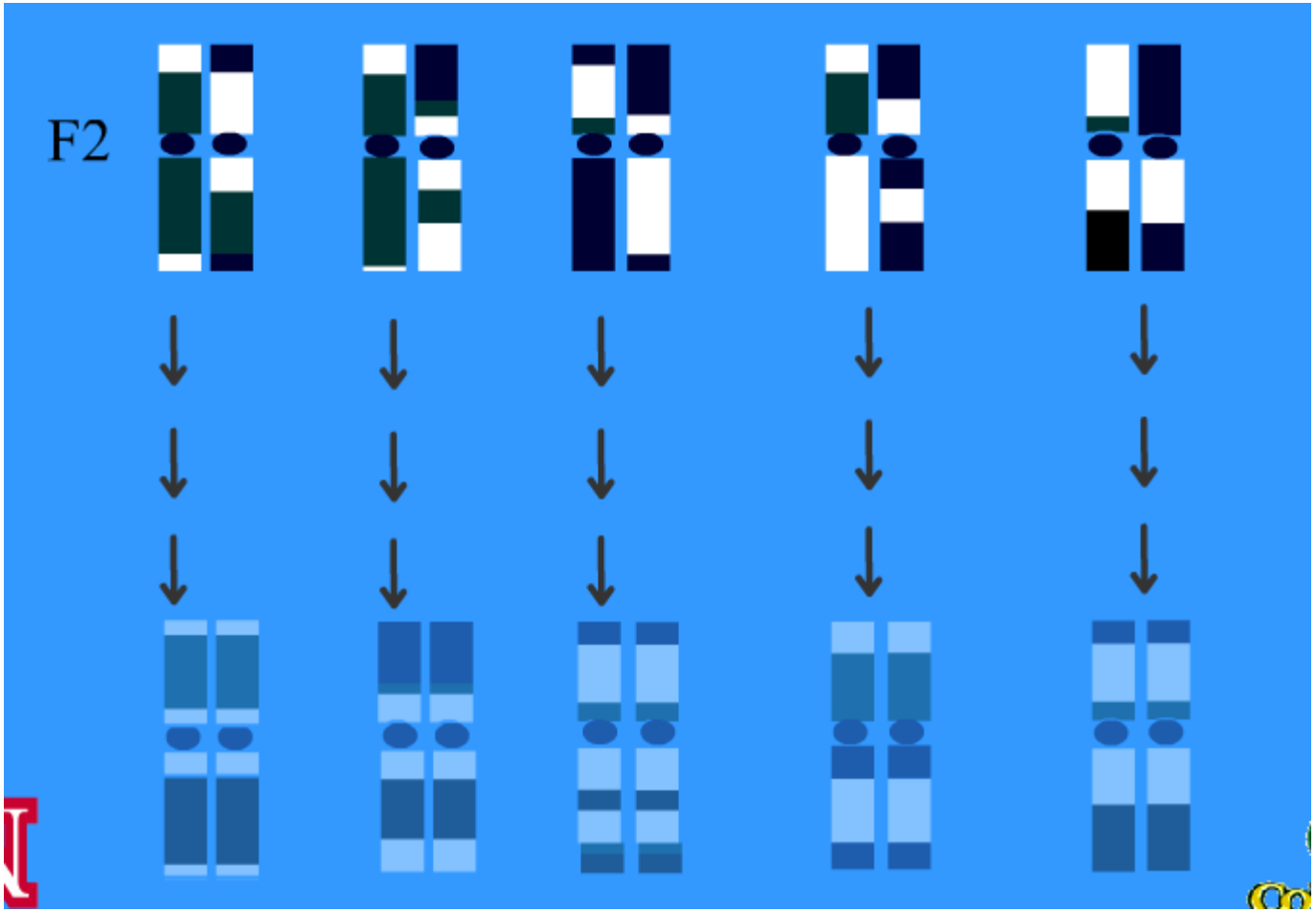
•Same lines can be used by multiple investigators - a standardized resource

•Good RI panels available for *Arabidopsis*, maize, *C. elegans*, *Drosophila*, and mice.

Another common QTL population is a set of recombinant inbred lines (RILs). RILs are developed by several generations of self-pollination from F2 plants, through a process known as **single-seed descent.**



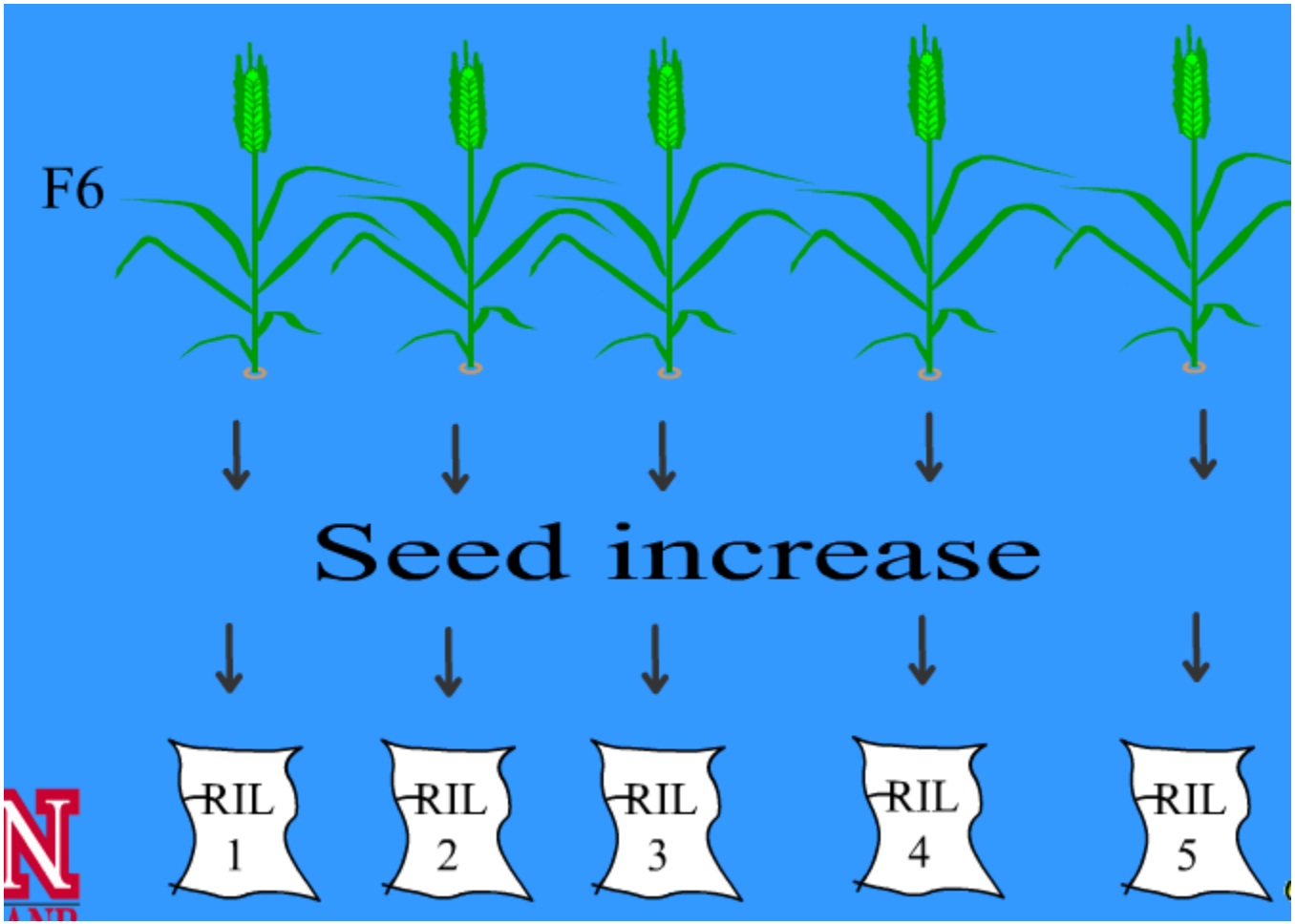
Because RILs have undergone more generations of reproduction, there are more opportunities for genetic recombination than in an F2 population.



By the F6 generation each line is nearly homozygous at all loci, as indicated in the diagram.

Each line is homozygous and homogeneous but tremendous variability is among the lines

The seed can then be multiplied to use for trait and DNA marker analysis.



Doubled haploid Lines(DHL)

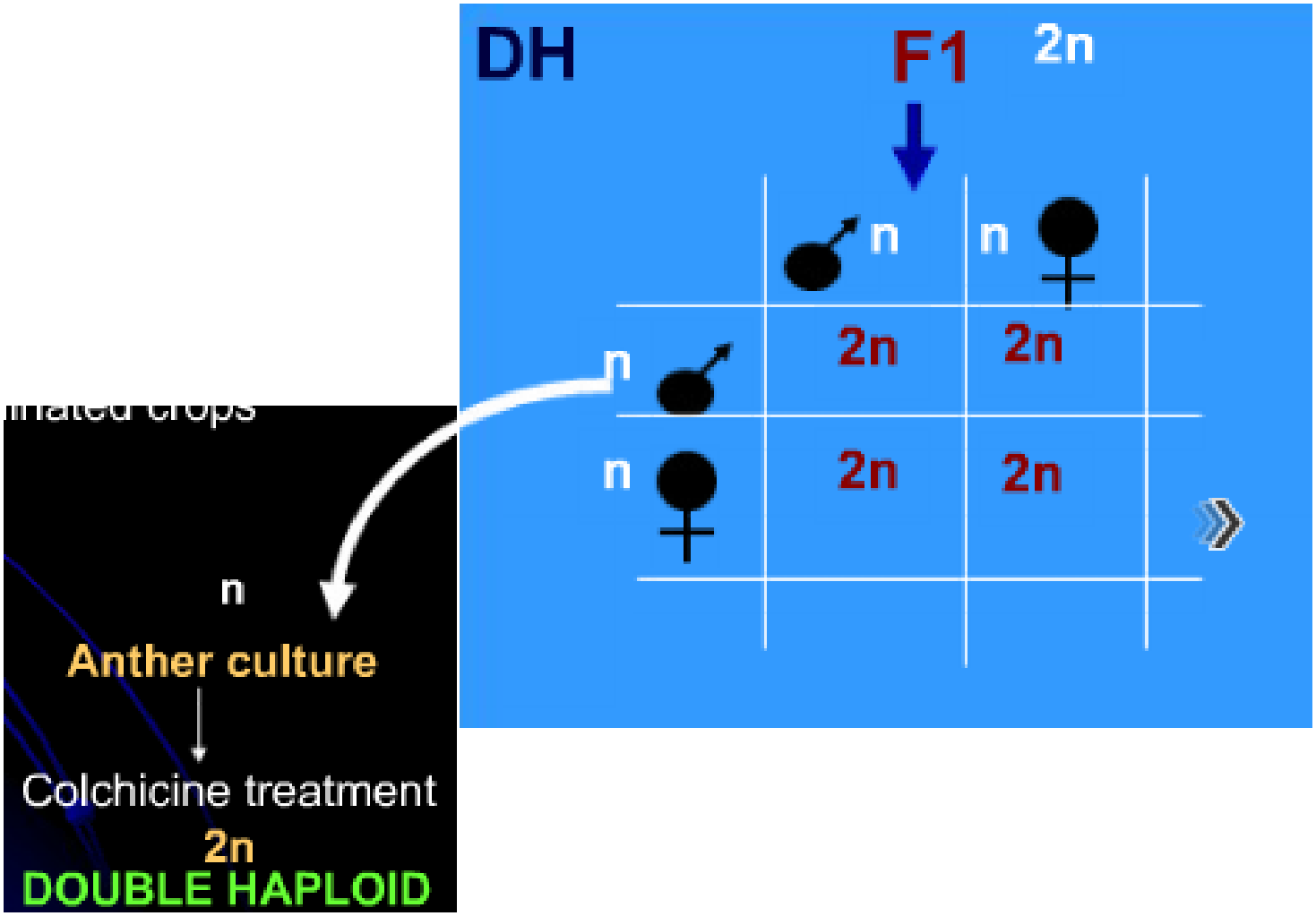
AA	aa	aa	AA	aa	AA
BB	bb	BB	bb	bb	BB
cc	CC	cc	cc	cc	CC
dd	DD	dd	dd	DD	dd
ee	ee	EE	EE	ee	ee
FF	ff	FF	FF	FF	ff

Advantages: 1)Spontaneous chromosome doubling of Haploid microspores in *in vitro* culture

2)Homozygosity achieved in a single step Plants.

Disadvantages: Less recombination between linked markers

Not all systems are amenable to in vitro culture

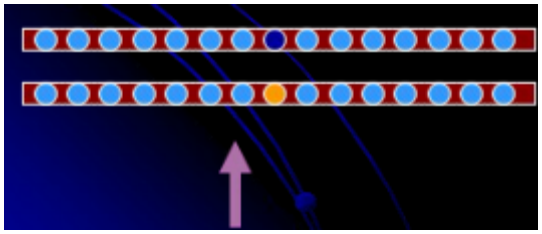

DH
F1
2n
n
n
2n
2n
n
n
2n
2n
n
Anther culture
Colchicine treatment
2n
DOUBLE HAPLOID

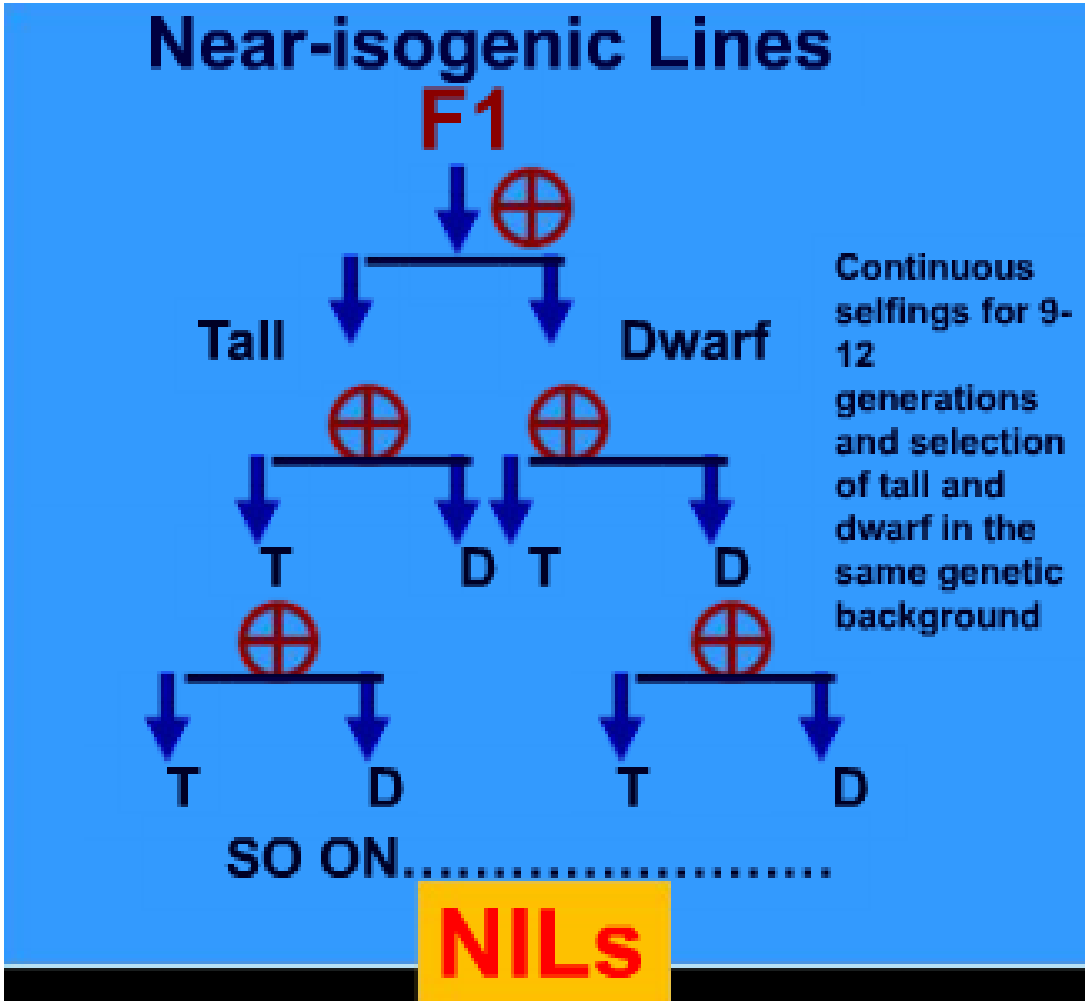
Near Isogenic Lines (NILs)

Advantage: Very precise and statistically strong, as background is constant; especially useful for validation experiments

Disadvantage: Can take time to construct; only useful for specific target QTL

Szalma SJ *et al.*



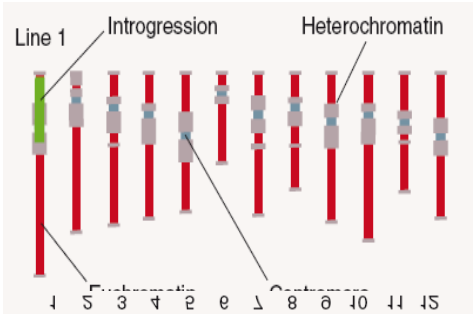

Near-isogenic Lines
F1
Tall
Dwarf
T
D
T
D
T
D
T
D
SO ON.....
NILs
Continuous selfings for 9-12 generations and selection of tall and dwarf in the same genetic background

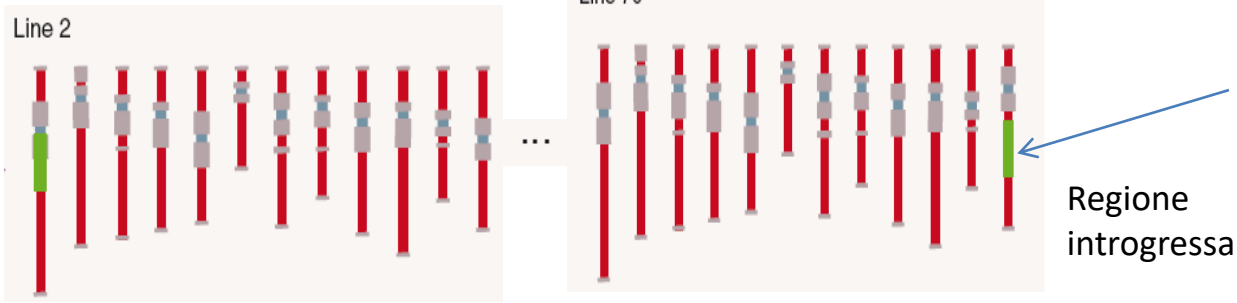
Introgression lines

ricorrente

donatore

Vantaggio: permettono una più semplice e sicura localizzazione dei QTL mascherando gli effetti epistatici di altri QTL presenti nel genoma di popolazioni sperimentali meno strutturate (es. BC, F_1, F_2…).





Regione introgressa

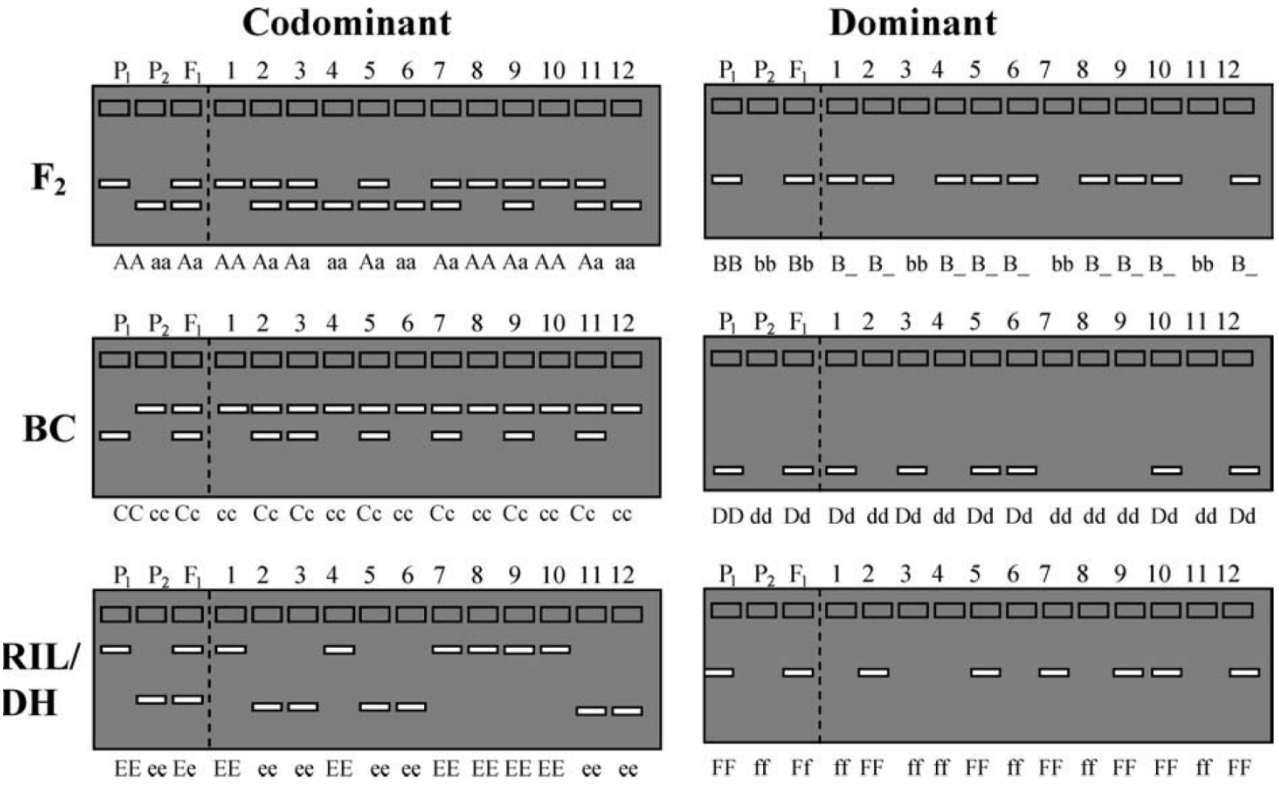
developed through multiple rounds of s | back-crossing (to the cultivated parent) between an elite cultivar, *Solanum lycopersicum* var. Roma, and a wild or 'ancestral' tomato plant, *Solanum pennellii*, to generate 76 independent introgression lines of tomato plants harboring chromosome segments from the wild relative[4]. Selection of specific, homozygous, single, overlapping chromosome introgressions in this population both simplifies QTL localization and defines linked DNA markers for use in crop improvement.

Identification of polymorphism
and
linkage analysis of markers

Identification of polymorphism

The second step in the construction of a linkage map is to identify DNA markers that reveal differences between parents (i.e. polymorphic markers). It is critical that sufficient polymorphism exists between parents in order to construct a linkage map (Young, 1994). In general, cross pollinating species possess higher levels of DNA polymorphism compared to inbreeding species.

Once polymorphic markers have been identified, they must be screened across the entire mapping population, including the parents (and F1 hybrid, if possible). This is known as marker 'genotyping' of the population. Therefore, DNA must be extracted from each individual of the mapping population when DNA markers are used (see examples next slide). The expected segregation ratios for codominant and dominant markers are presented in Table 2.



Significant deviations from expected ratios can be analysed using chi-square tests. Generally, markers will segregate in a Mendelian fashion

Table 2. Expected segregation ratios for markers in different population types

Population type	Codominant markers	Dominant markers
F_2	1: 2:1 (AA:Aa:aa)	3:1 (B_:bb)
Backcross	1:1 (Cc:cc)	1:1 (Dd:dd)
Recombinant inbred or doubled haploid	1:1 (EE: ee)	1:1 (FF:ff)

In some polyploid species such as sugarcane, identifying polymorphic markers is more complicated (Ripol et al., 1999). The mapping of diploid relatives of polyploid species can be of great benefit in developing maps for polyploid species. However, diploid relatives do not exist for all polyploid species (Ripol et al., 1999; Wu et al., 1992). A general method for the mapping of polyploid species is based on the use of single-dose restriction fragments.

Linkage analysis of markers

The final step of the construction of a linkage map involves coding data for each DNA marker on each individual of a population and conducting linkage analysis using computer programs. Missing marker data can also be accepted by mapping programs. Linkage between markers is usually calculated using odds ratios (i.e. the ratio of linkage versus no linkage). This ratio is more conveniently expressed as the logarithm of the ratio, and is called a logarithm of odds (LOD) value or LOD score (Risch, 1992). LOD values of >3 are typically used to construct linkage maps. A LOD value of 3 between two markers indicates that linkage is 1000 times more likely (i.e. 1000:1) than no linkage (null hypothesis).

LOD values may be lowered in order to detect a greater level of linkage or to place additional markers within maps constructed at higher LOD values.

Commonly used software programs include Mapmaker/ EXP (Lander et al., 1987; Lincoln et al., 1993a) and MapManager QTX (Manly et al., 2001), which are freely available from the internet. JoinMap is another commonly-used program for constructing linkage maps (Stam, 1993).

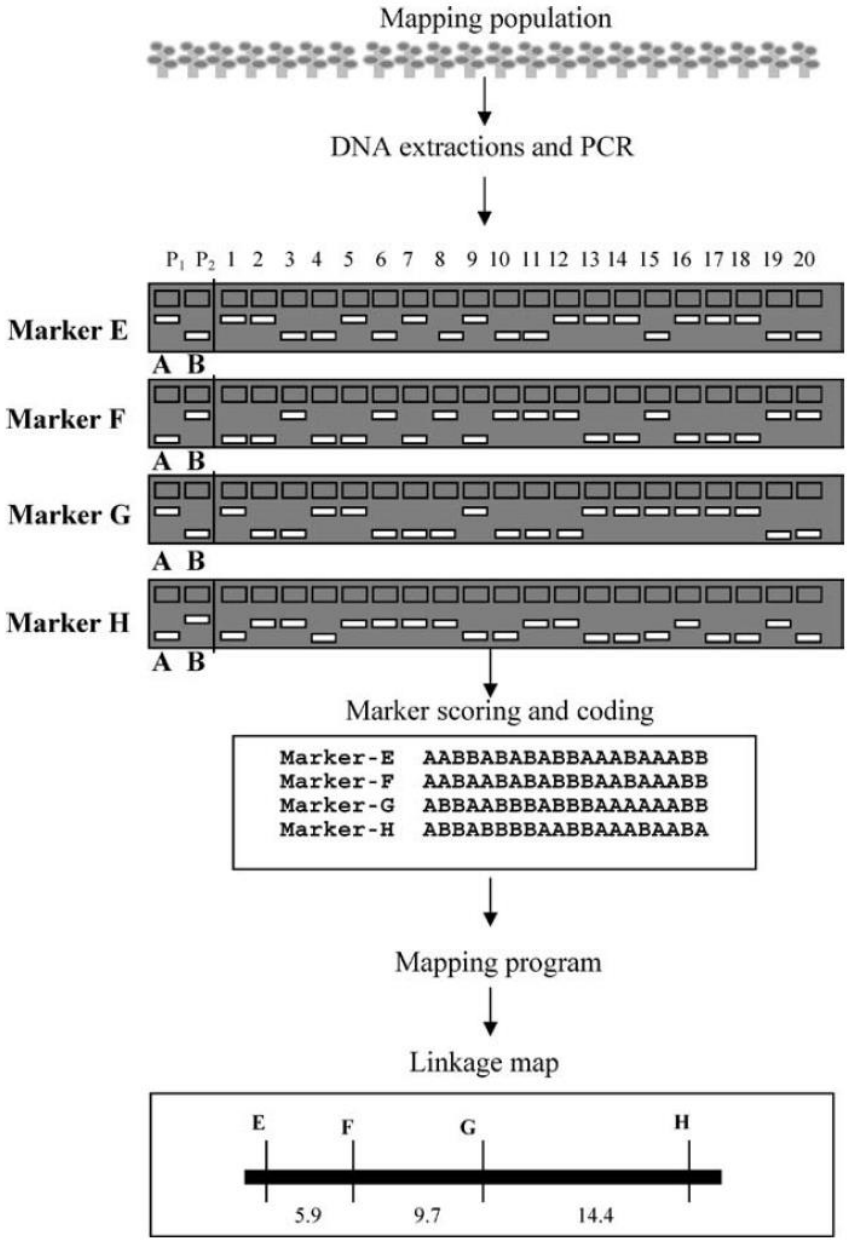


Figure 6. Construction of a linkage map based on a small recombinant inbred population (20 individuals). The first parent (P_1) is scored as an 'A' whereas the second parent (P_2) is scored as a 'B'. Coding of marker data varies depending on the type of population used. This linkage map was constructed using Map Manager QTX (Manly et al., 2001) using the Haldane mapping function.

A typical output of a linkage map.

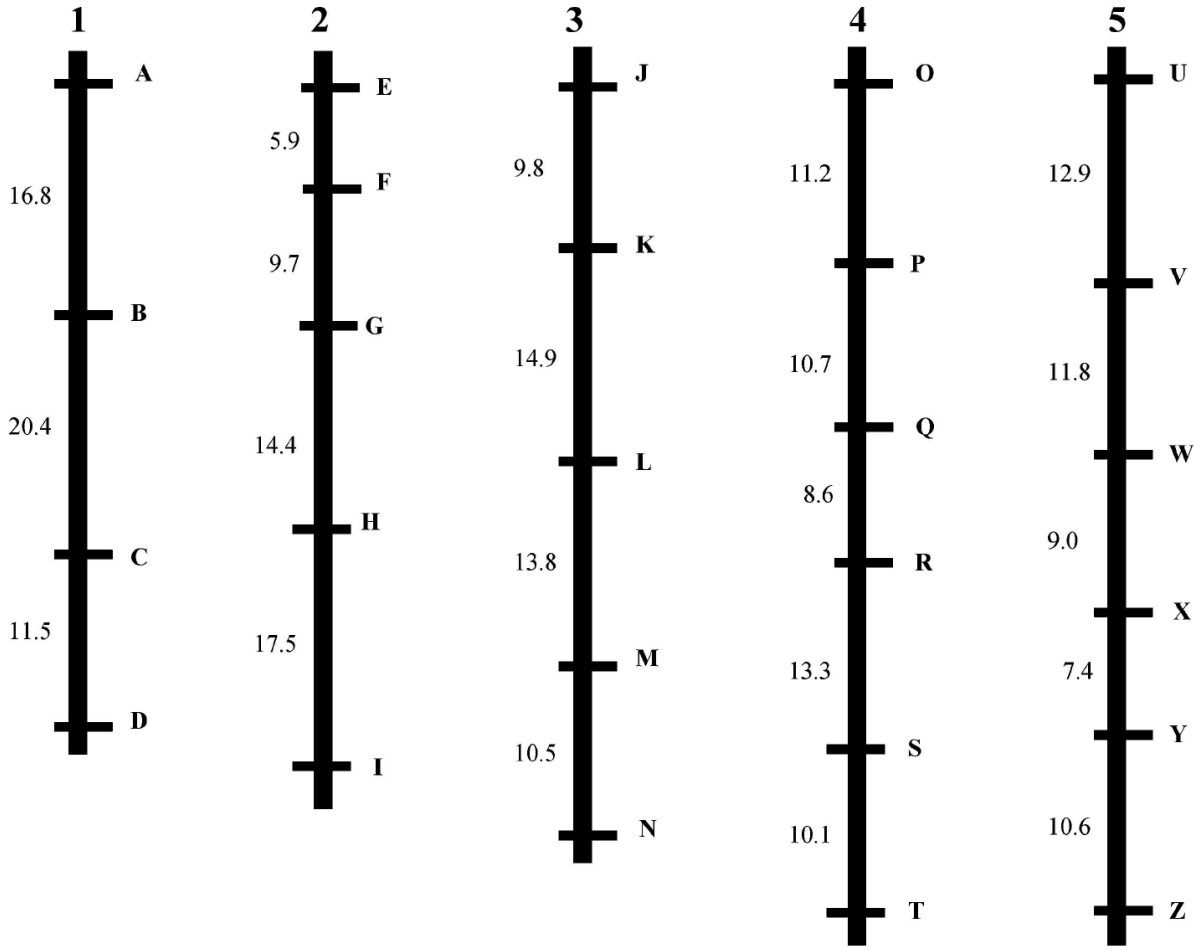


Figure 7. Hypothetical 'framework' linkage map of five chromosomes (represented by linkage groups) and 26 markers. Ideally, a framework map should consist of evenly spaced markers for subsequent , analysis.

Linked markers are grouped together into 'linkage groups,' which represent chromosomal segments or entire chromosomes

A difficulty associated with obtaining an equal number of linkage groups and chromosomes is that the polymorphic markers detected are not necessarily evenly distributed. over the chromosome, but clustered in some regions and absent in others (Paterson, 1996a). In addition to the non-random distribution of markers, the frequency of recombination is not equal along chromosomes.

The accuracy of measuring the genetic distance and determining marker order is directly related to the number of individuals studied in the mapping population.
Ideally, mapping populations should consist of a minimum of 50 individuals for constructing linkage maps.

Mapping functions are required to convert recombination fractions into centiMorgans (cM) because recombination frequency and the frequency of crossing-over are not linearly related (Hartl & Jones, 2001;Kearsey&Pooni, 1996). When map distances are small (<10 cM), the map distance equals the recombination frequency. However, this relationship does not apply for map distances that are greater than 10 cM (Hartl & Jones, 2001). Two commonly used mapping functions are the Kosambi mapping function, which assumes that recombination events influence the occurrence of adjacent recombination events, and the Haldane mapping function, which assumes no interference between crossover events

It should be noted that distance on a linkage map is not directly related to the physical distance of DNA between genetic markers, but depends on the genome size of the plant species (Paterson, 1996a).
Furthermore, the relationship between genetic and physical distance varies along a chromosome (Kunzel et al., 2000; Tanksley et al., 1992;Young, 1994). For example, there are recombination ‘hot spots’ and ‘cold spots,’ which are chromosomal regions in which recombination occurs more frequently or less frequently, respectively.

Mapping with a high resolution requires a high density of genetic markers

•Several Arabidopsis accessions, or ecotypes, are sufficiently divergent to support the design of molecular markers at this high density.
•The most commonly used combination for mapping purposes is Landsberg *erecta* X Columbia (Ler X Col).
•These two accessions have been estimated to differ in 4 to 11 positions every 1,000 bp (Chang et al., 1988_ Konieczny and Ausubel, 1993_ Hardtke et al., 1996_.
•Most existing mutations, including mutations causing visible phenotypes that can be employed as genetic markers have been induced either in a Col or a Ler background.

Molecular markers commonly used for mapping (SSRs, CAPS and dCAPS)

•codominant (both chromosomes of a plant may be genotyped)
• PCR-based and can be analyzed on agarose gels (easy to use and inexpensive)
•Several methods for the extraction of plant DNA from small tissue samples have been described
•These methods are cheap and suitable for high throughput applications in microtiter plates.