Genetic linkage map

Bioinformatics Applications (PLPTH813)

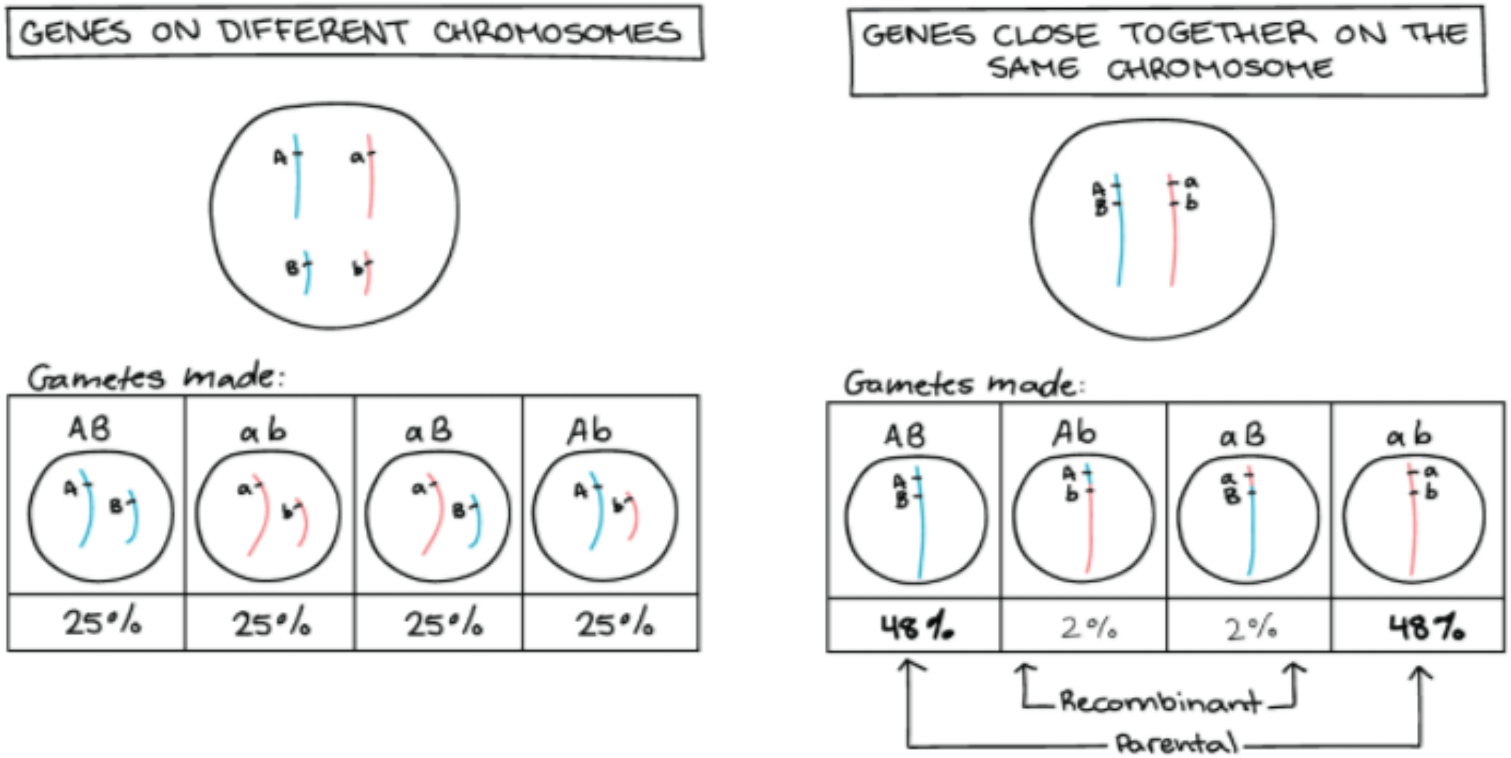
Ying Hu

03/07/2017

Outline

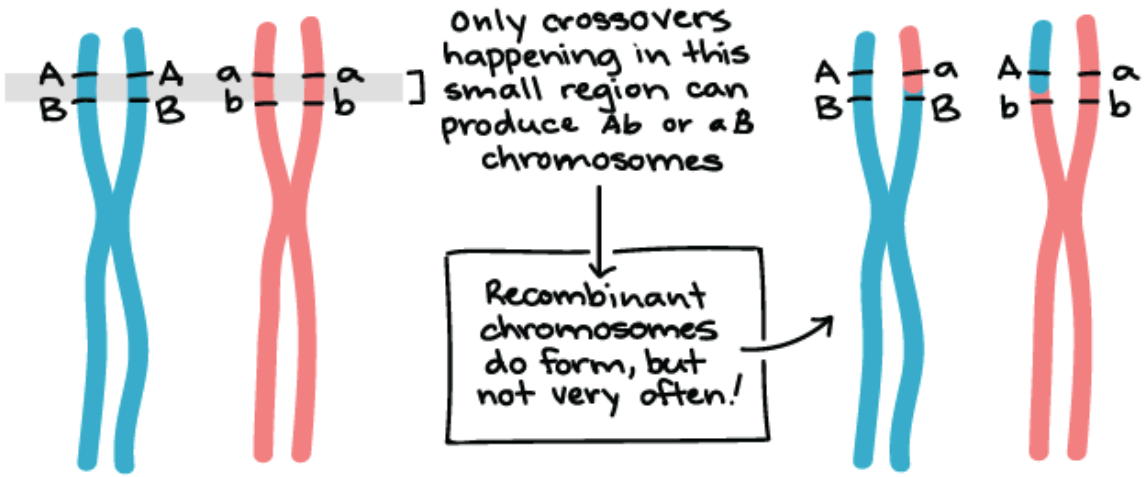
1. Introduction of Genetic linkage maps
2. Construction of Genetic linkage maps
3. An example of constructing a genetic map using qtl/r

What is genetic linkage?



When genes are linked, genetic crosses involving those genes will lead to ratios of gametes (egg and sperm) and offspring types that are not what we'd predict from Mendel's law of independent assortment.
This phenomenon is called **genetic linkage**.

Finding recombination frequency



- The frequency of recombination events between two genes (i.e., their degree of genetic linkage) is used to estimate their relative distance apart on the chromosome.
- Two very close-together genes will have very few recombination events and be tightly linked, while two genes that are slightly further apart will have more recombination events and be less tightly linked.

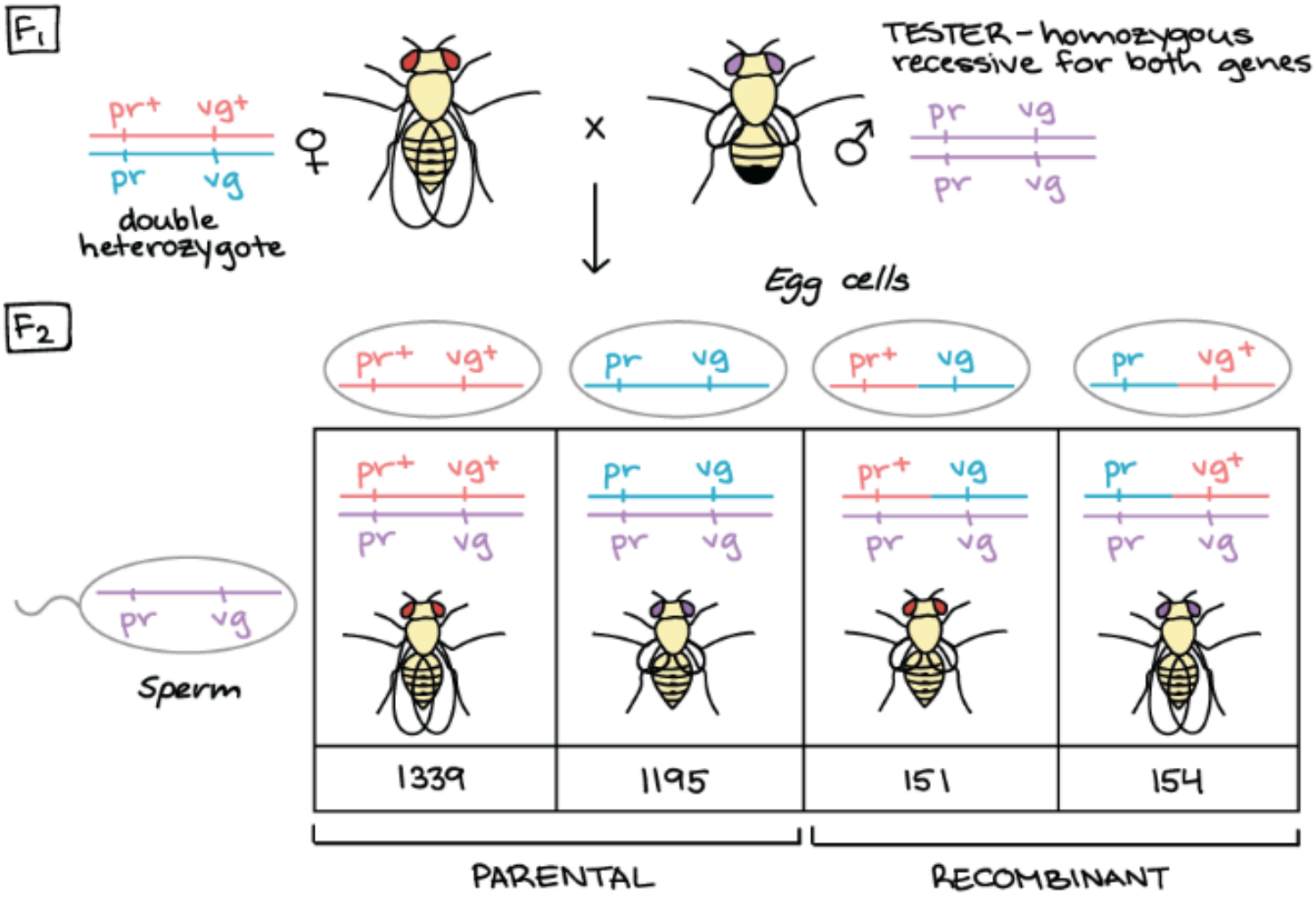
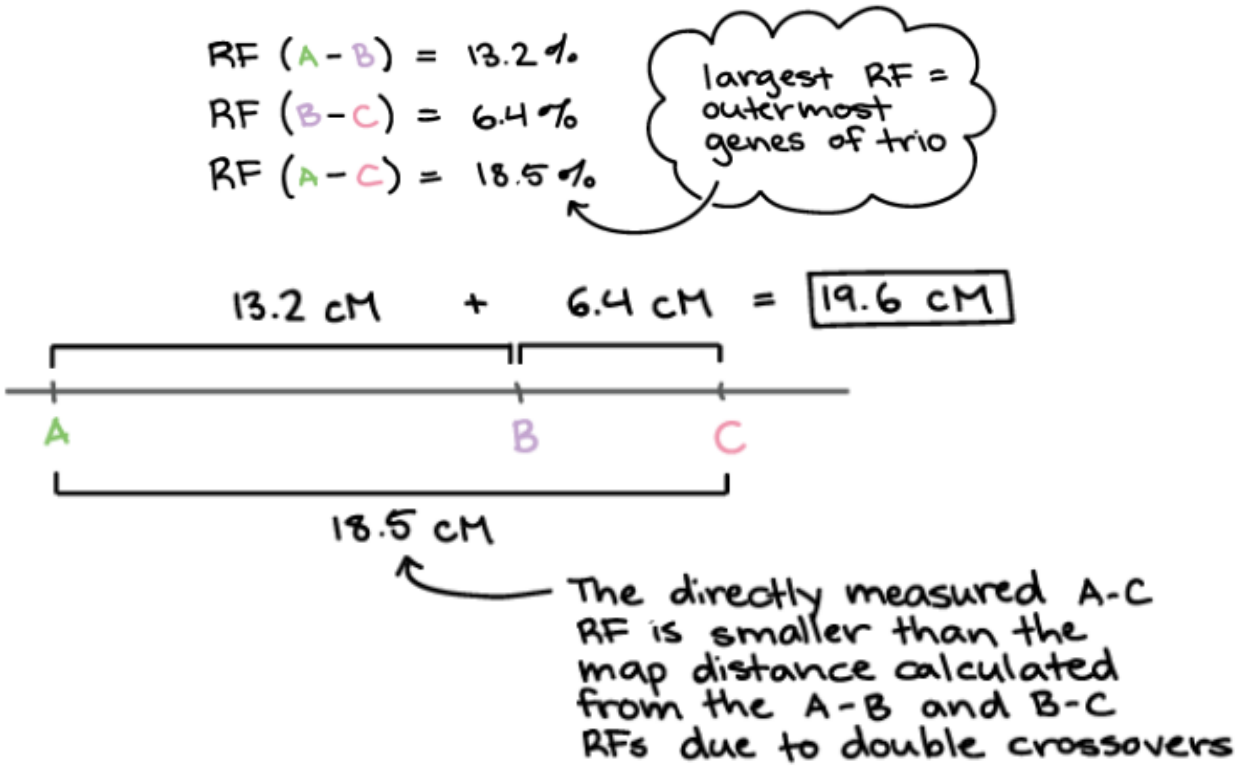


Image modified from "Drosophila melanogaster," by Madboy74 (CC0/public domain).

$$\mathrm{RF} = \frac{151 + 154}{1339 + 1195 + 151 + 154} \times 100\% = 10.7\%$$

Recombination frequencies can be used to map the order of genes and the relative distances between genes (loci) on linear linkage groups.

1 map unit = 1% recombination = 1 centimorgan



Genetic linkage maps

- Genetic linkage mapping dates back to the early 20th century when scientists began to understand the recombinational nature and cellular behavior of chromosomes.
- During the past few decades the introduction of DNA-based markers such as RFLPs, RAPDs, SSRs and AFLPs caused genetic maps to become much more densely populated.
- Most recently, the accumulation of sequence information has led to a further leap in marker density, principally driven by very high throughput and highly accurate next generation sequencing technology.

A genetic map = a linear ordering of markers (also called *loci*) along the linkage group(chromosome).

Linkage group = group of genes, each of which is linked ($r < 0.5$) to at least one other.

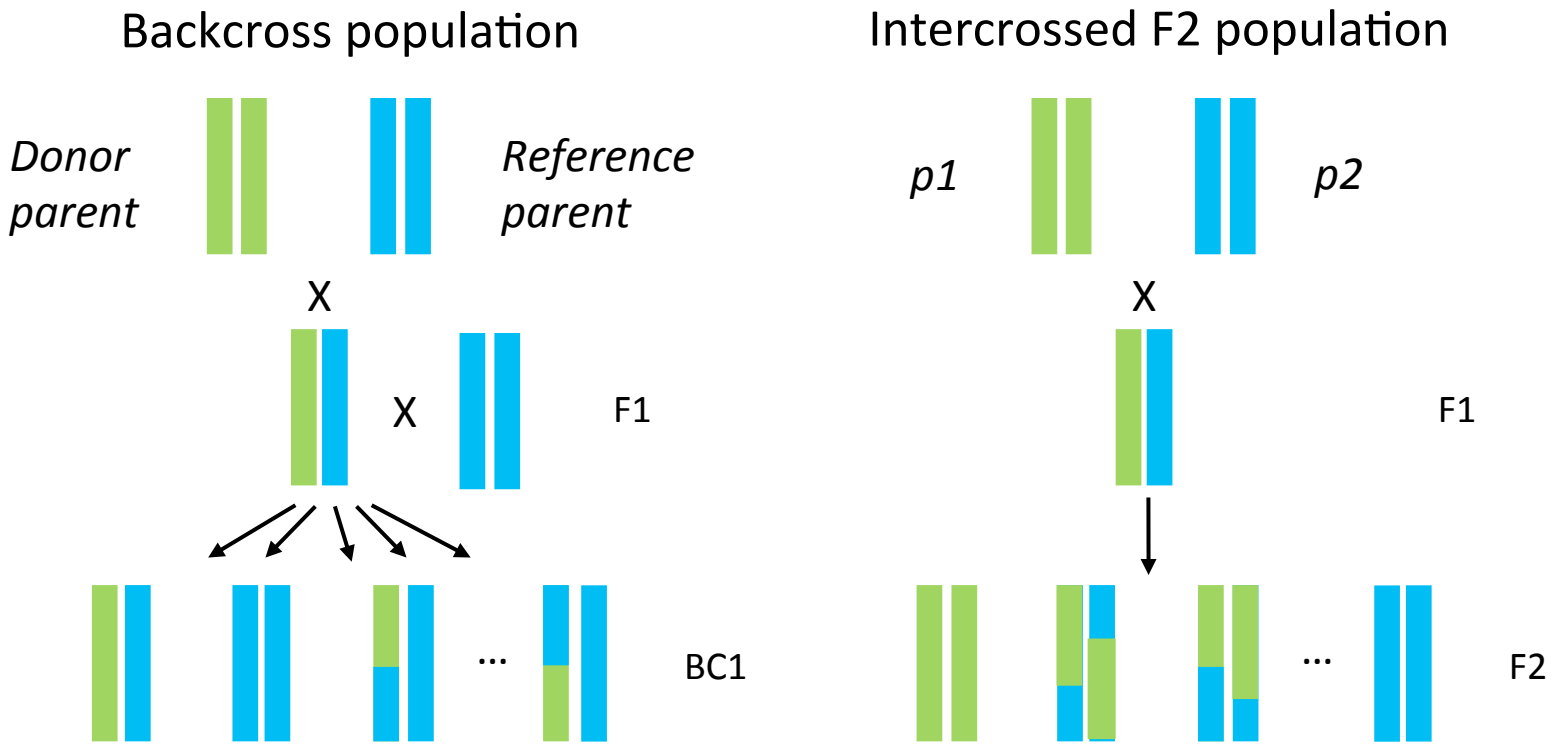
Construction of Genetic linkage maps

1. Develop appropriate mapping population and decide the sample size.
2. Decide the type of molecular marker(s) for genotyping the mapping population.
3. Screen parents for marker polymorphism, and then genotype the mapping population (parents plus all progenies).
4. Perform linkage analyses (calculate pairwise recombination frequencies between markers, establish linkage groups, estimate map distances, and determine map order) using statistical programs.

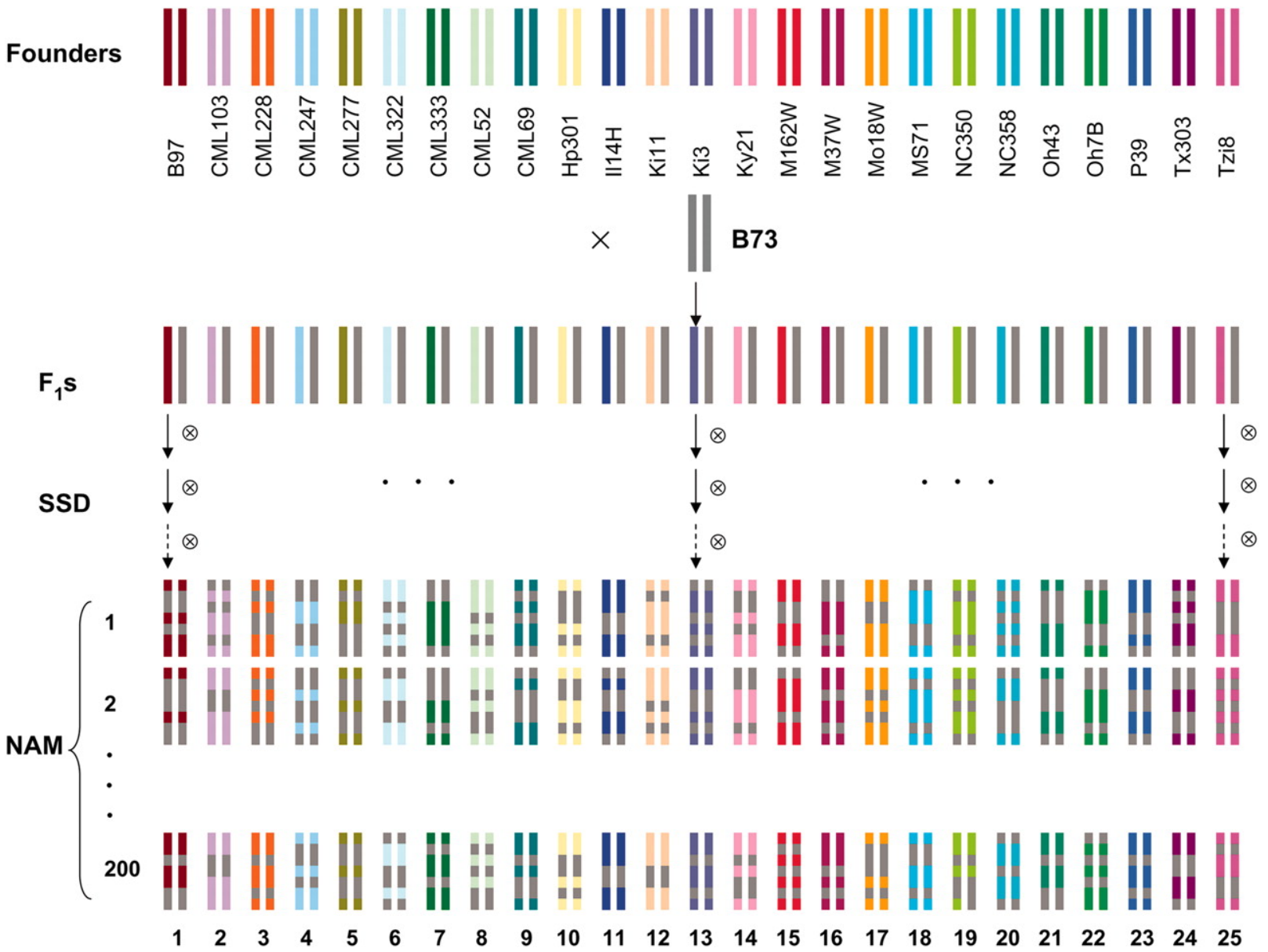
Construction of Genetic linkage maps

Population used for constructing genetic linkage maps

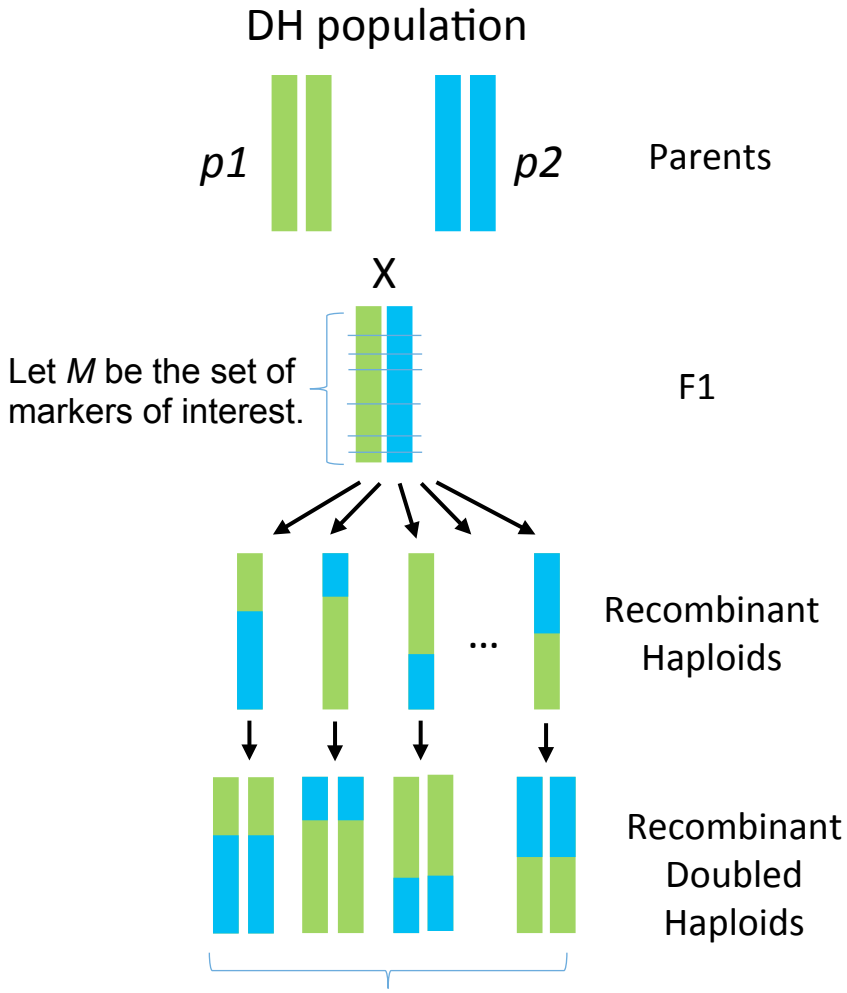
- *BackCross* (BC1)
- *Doubled Haploid* (DH)
- *Recombinant Inbred Line* (RIL)
- *Intercrossed F2* (F2)


Backcross population
Donor parent
Reference parent
X
X
F1
...
BC1
Intercrossed F2 population
p1
p2
X
F1
...
F2

NAM recombinant inbred lines (RILs)



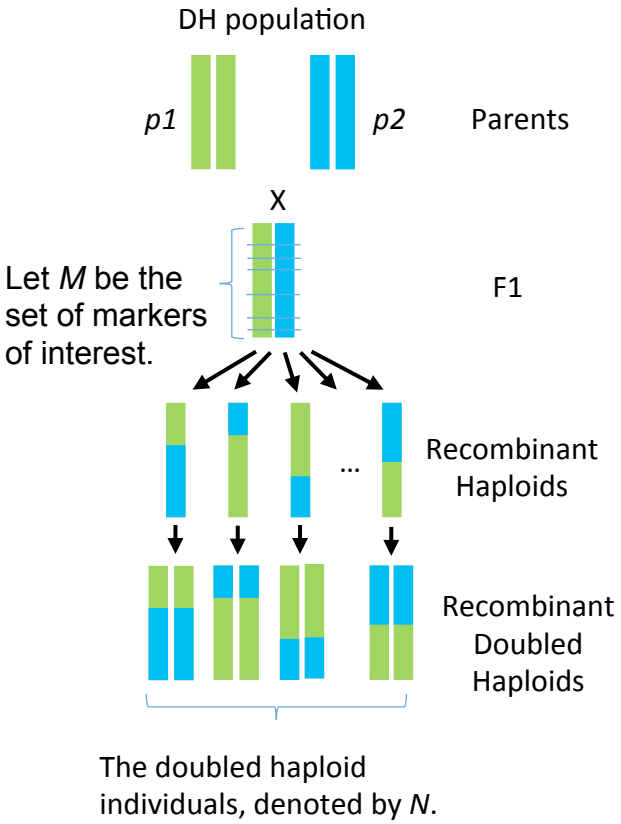
Construction of Genetic linkage maps



The doubled haploid individuals, denoted by *N*.

Assumption:

- The parents p1 and p2 are homozygous on every markers.
- The same marker always has different allelic states in the two parents.



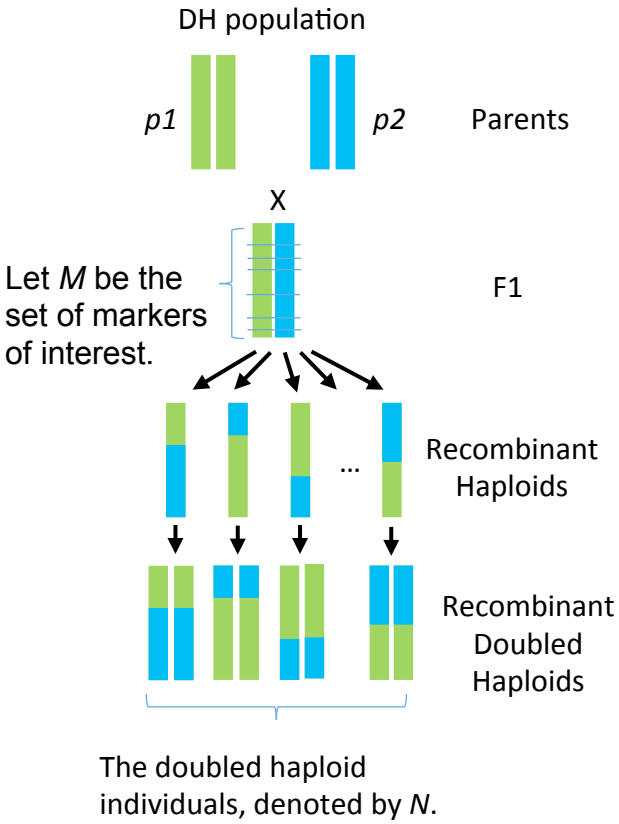
$$\mathbb{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M1} & \cdots & a_{MN} \end{bmatrix}$$

Individuals →

Markers ↓

Given a marker $l_i \in M$, we use $\mathbb{A}[i,]$ to refer to the row corresponding to l_i.

Given an individual $c_k \in N$, we use $\mathbb{A}[,k]$ to refer to the column corresponding to c_k.



$$\mathbb{A} = \begin{matrix} l_1 \\ l_2 \\ \\ \\ \\ \end{matrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M1} & \cdots & a_{MN} \end{bmatrix}$$

Individuals →

↓ Markers

Two steps to build a genetic linkage map from the matrix $\mathbb{A}$.

1. Determine markers from the same linkage group (chromosome).

For a pair of markers $l_1, l_2 \in$ M and an individual c $\in$ N

a. If c has genotype A on l_1 and genotype B on l_2 (or vice versa), c is a recombinant with respect to l_1 and l_2.

b. $0 < P_{i,j} < 0.5$ ($P_{i,j}$ the probability of a recombinant event with respect to a pair of markers (l_1, l_2))

2. Determine the correct order of markers on the chromosome.

Calculate pairwise recombination frequencies between markers, estimate map distances, and determine map order

Construction of Genetic linkage maps: R/qtl

QTL analysis package R/qtl (Broman & Wu, 2014) is a very popular package for the linkage map construction of a simple Backcross (BC), Doubled Haploid (DH), intercrossed F2 (F2), 4-way crosses and advanced Recombinant Inbred Lines (RIL).