Trait/DNA-marker associations

(in polyploids)

Isabel Roldán-Ruiz

August 2012

Institute for Agricultural and Fisheries Research
Plant Sciences Unit
www.ilvo.vlaanderen.be
Agriculture and Fisheries Policy Area

Contents

1. Defining the target
2. Identifying suitable germplasm
3. Mapping population structures
4. Phenotypic evaluation
5. *Genotyping (as discussed in lecture 1)*
6. Strategies to identifying marker/trait associations
   - Linkage analysis (in polyploids)
   - Bulked Segregant Analysis
   - QTL analysis
   - Association Mapping
7. Developing markers for application – marker validation

Only at this step the markers are tested in breeding plant materials

Introduction

Remember from Lecture 1:

- DNA-marker technologies offer the possibility to screen plant genomes for sequence polymorphisms
- DNA-marker polymorphisms help us to define specific positions (loci) in the genome
  - ⇒ DNA-markers make it possible to follow chromosome pieces from one generation to the next
- How to identify marker/trait associations? How do we identify DNA-markers which are located **in or in the neighborhood of** the genes of interest?
  - In most cases => construction, phenotyping and genotyping (using DNA-markers) of dedicated populations (= mapping populations)
  - But other possibilities exist; e.g. association mapping

ILVO

1. Defining the target

- Is the trait of importance to breeding program or to biological research?
- Is a DNA-marker needed?
  - What is the cost of a bioassay relative to a marker assay?
    - Financial aspects / time constraints (can we save one generation of crossing by using DNA-markers?)
  - Is the trait dominant or recessive?
    - Recessive traits may be hard to identify in a bioassay => DNA-marker development is then advantageous
  - Perhaps no alternative to marker development
    - Quarantaine trait (the disease is not present in the country in which breeding is carried out) => no bioassay possible
    - Pyramiding disease resistances to avoid resistance breakdown
    - Map-based cloning of genes – high resolution map required
    - Gene deployment – where desirable alleles are available for several loci, but only one is really needed. How does one decide on the best one to use?

2. Identifying suitable germplasm

CHOICE OF CONTRASTING PARENTS

We can only find markers linked to a trait of interest if we compare plants with and without the trait

- Among available germplasm search for plants with and without the trait
- Or find an existing mapping population that segregates for the trait
- Remember: only traits that segregate in the population can be mapped

Example: to identify genomic loci linked to plant height, which plants would you select as parents of the mapping population? Could you also map seed production in this mapping population? And disease resistance?

	height	seed production	disease resistance
Plant 1	100	200	++++
Plant 2	105	210	++++
Plant 3	150	205	+
Plant 4	95	200	+++

3. Population structures

DESIGN OF STUDY

- Knowledge on genetics
  - Monogenic / multigenic trait
  - Heritability

  Marker development is specially advantageous for complex traits with low heritability in the global population (not specifically in the population used for mapping)
- Decisions about population structure
  - Double haploids (DH)
  - F_2
  - Recombinant inbred lines (RILs) created by single seed descent
  - Near Isogenic Lines (NILs) created by back-crosssing
  - Pair-cross between highly heterozygous parents (CP – cross pollinator- population)
- Population size (how many progeny plants do we need?)
  - The higher the population size the higher the 'resolution' of the mapping study
    - To determine chromosome location of a single gene => 50 F_2 is appropriate
    - Map-based cloning => over 1000 progeny plants required
    - Numbers depend on genome size and complexity

Sweetpotato - only one possibility: pair-cross between heterozygous parents; many progeny plants

ILVO

4. Phenotypic characterization

- Phenotypic evaluation should be possible for single plants
- Quality, yield and traits of low heritability (strong environmental influences) => many replicates and /or multiple field trials
- Association mapping => not based on mapping populations; phenotypic information can be collected from existing programs

Qualitative and quantitative traits

DEFINITION

- From the point of view of the molecular geneticist (interested in the identification of DNA-markers linked to genes involved in agronomic traits) there is one question of special relevance:
  - Some traits (e.g. incompatibility, some disease resistances) are controlled by one or two genes with large effects
    - Differences in expression (phenotype) are '**qualitative**'
    - You can say that the traits follow a 'Mendelian' inheritance
  - Most traits of agronomic relevance such as growth or yield are complex, being under the control of a number of genes as well as the environment
    - Differences in expression (phenotype) are '**quantitative**'. No clear discontinuity exists between the phenotype of different genotypes, giving the impression of a continuous distribution

Knowledge about the kind of trait we are dealing with is of crucial relevance for the design of the experimental strategy

Field for phenotypic evaluation of a mapping family

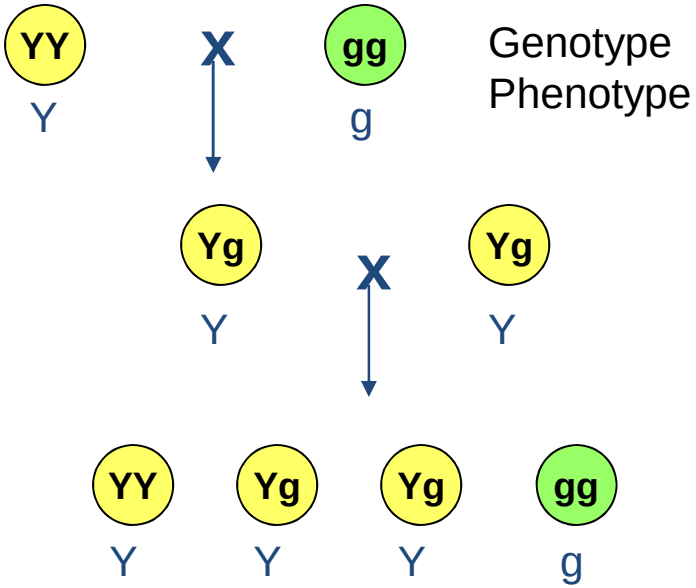
2808 wild *Malus* seedlings from 310 populations from Kazakhstan, Russia, China & Turkey
Phil Fosline, PGRU, USA

Qualitative and quantitative traits

DEFINITION

Monogenic trait

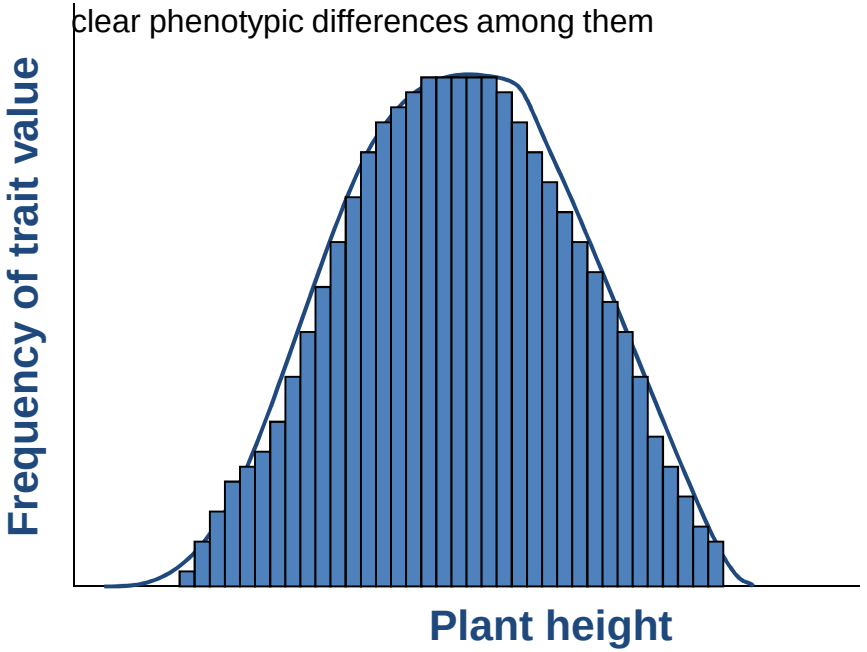
- Remember: Mendel studied major gene differences in the garden pea: color of the seed (one locus with alleles Y and g, Y>g)
- But: some factors can affect the expression of Mendel's laws



If we analyze the diversity in a collection of plants for this character, we are only looking at the genetic diversity contained in this genetic locus

Multigenic trait

- Multiple genes (and environment) affect the expression of the trait
- The expression in the population is a 'bell-shaped' curve: there are many genotypes and there are no clear phenotypic differences among them

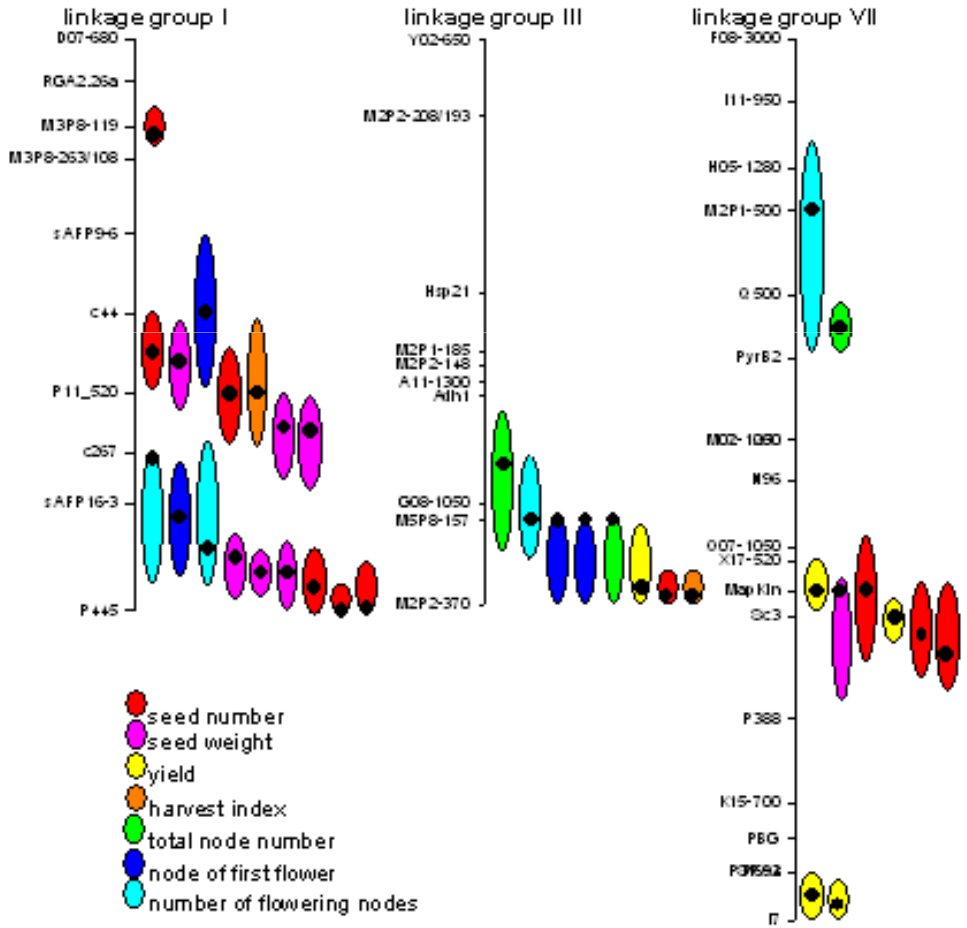


If we analyze the diversity in a collection of plants for this character, we are looking at the genetic diversity contained in all the genetic loci which influence the trait

Qualitative traits

- Mendel studied major gene differences in the garden pea: colour of the seed (one locus with alleles Y and g, Y>g)
- But: some factors can affect the expression of Mendel's laws
  - Environmental influences can complicate the interpretation of the results
  - Presence of lethal alleles (homozygotes die at early developmental stage)
  - Presence of multiple alleles (more than two alleles at a single locus in the population)
  - Incomplete dominance (heterozygous individuals show an intermediate phenotype) and co-dominance (heterozygous individuals show a distinct phenotype)

Quantitative traits



Linkage mapping of QTLs for seed yield, yield components and developmental traits in pea (*Pisum sativum* L.)
Timmerman-Vaughan et al. (2004) New directions for a diverse planet: Proceedings of the 4th International Crop Science Congress Brisbane, Australia, 26 Sep – 1 Oct 2004 | ISBN 1 920842 20 9 | www.cropscience.org.au

If we analyze the diversity of seed size in a collection of pea plants, we are simultaneous looking at the genetic diversity contained in several genetic loci

Quantitative traits

DIFFICULTIES

- <u>Different genotypes can display the same phenotype</u>
  - o 3 genes each with small equal effect on the phenotype
  - o each gene has two co-dominant alleles
  - o $3^3 = 27$ *genotypes* are possible, but we only distinguish 7 *phenotypes*

	genotype	**phenotype**
1.	AABBCC	6
2.	AABBCc	5
3.	AABBcc	4
4.	AABbcc	3
5.	AaBbCc	3
6.	AAbbCc	3
7.	AAbbCC	4
…		
27.	Aabbcc	0

Three different genotypes give the same phenotype

- <u>Dominance can vary from gene to gene</u>
  - o 3 genes: 2 co-dominant, 1 dominant

- <u>Environmental variation:</u> the same genotype can display different phenotypes depending on the environment
  - o Genotype and environment might interact obscuring genetic effects

Quantitative traits

DIFFICULTIES

- Epistasis
  - Non-allelic interaction: the effect of the genotype at one locus depends on the genotype at another locus

No epistasis

	aa	Aa	AA	Genotypic difference AA - aa
bb	1	2	3	3 – 1 = 2
Bb	2	3	4	4 – 2 = 2
BB	3	4	5	5 – 3 = 2

The change in phenotype when AA is replaced by aa is the same whatever the genotype at the B/b locus

Epistasis

	aa	Aa	AA	Genotypic difference AA - aa
bb	1	4	9	9 – 1 = 8
Bb	4	9	16	16 – 4 = 12
BB	9	16	25	25 – 9 = 16

The change in phenotype when AA is replaced by aa varies considerable depending on the B/b genotype

6. Strategies to identify marker/trait associations

STRATEGIES

Qualitative traits

Linkage map construction + mapping of the trait

- The chromosomal location of a ‘phenotype’ or a ‘mutation’ is determined by identifying nearby genetic markers which are co-transmitted from parent to progeny with the phenotype
- Requires extensive phenotyping and genotyping of many plants of a mapping (=segregating) population
- Typical experiment: parents and >100 offspring plants + hundreds to thousands of molecular markers

Bulk Segregant Analysis (=BSA)

- Allows to reduce the amount of genotyping work
- Typical experiment: parents and 2 bulks + hundreds to thousands of molecular markers

Quantitative traits

Quantitative Trait Loci-mapping (=QTL-mapping)

- Requires a detailed linkage map (hundreds of molecular markers), using large progenies (typically 200 or more)
- Requires sophisticated statistical tools

Linkage Disequilibrium mapping (for quantitative and qualitative traits)

- Uses ‘natural’ or breeding populations to map traits by means of association analysis
- Mostly used in humans, recently extended to plants
- Requires sophisticated statistical tools; extremely difficult in highly heterozygous crops

Linkage mapping

Institute for Agricultural and Fisheries Research
Plant Sciences Unit
www.ilvo.vlaanderen.be
Agriculture and Fisheries Policy Area

Linkage mapping

EXAMPLE

Two phenotypic traits segregating in a progeny

In corn, colored aleurone (in the kernel) is due to one dominant allele *R*. The recessive allele (of the same locus) *r*, produces colorless aleurone. The plant color is controlled by another locus, with two alleles (*Y* and *y*). *Y* is dominant and results in green color, while *y* results in yellow color. In a cross between a plant with colored aleurone and green color, and a plant that is homozygous recessive for both traits, the following progeny was obtained:

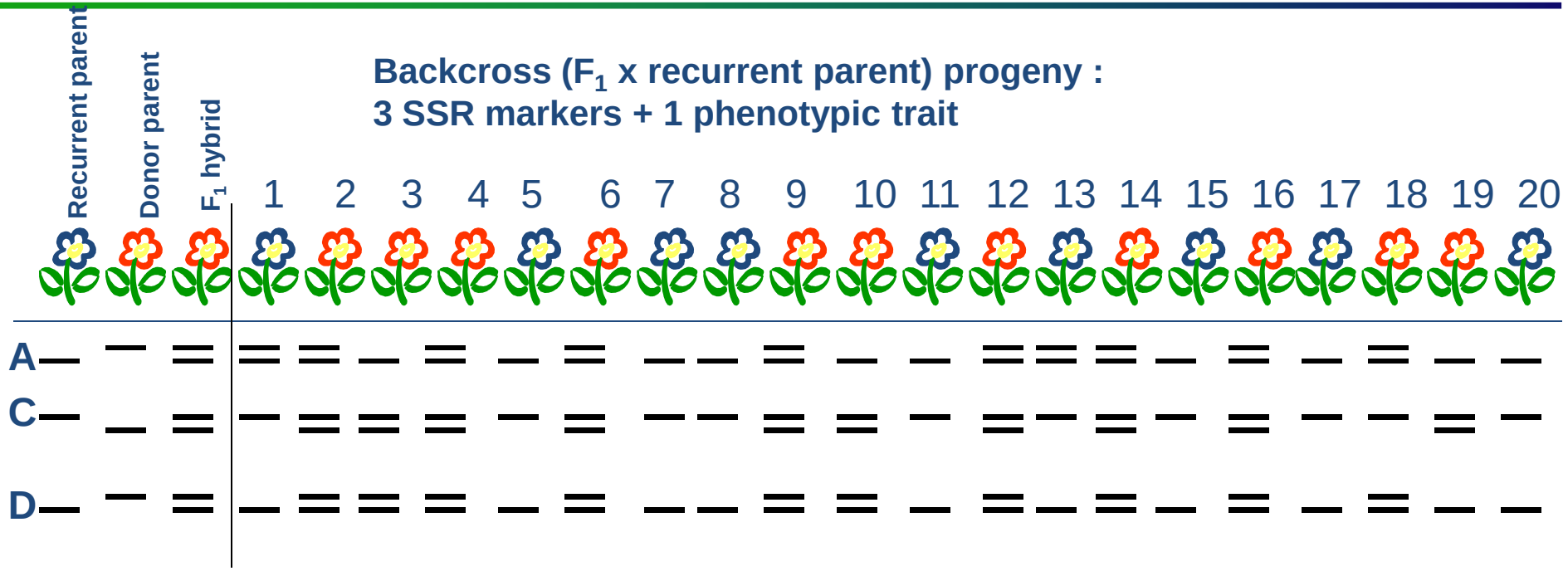
Colored + green	88	**88 R_Y_**
Colored + yellow	12	**12 R_yy**
Colorless + green	8	**8 rrY_**
Colorless + yellow	92	**92 rryy**

Parent plant: RrYy or RryY or RRYY

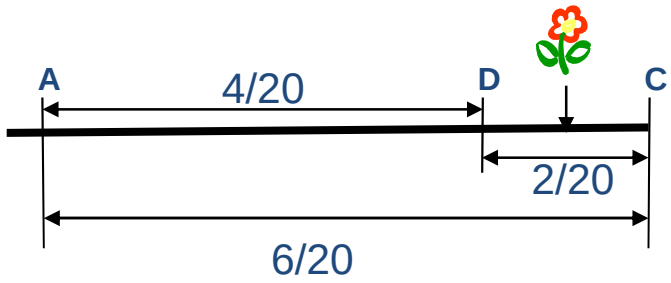
RY (parental)	**ry = RrYy**	**88**	**(12+8)/(88+12+8+92) = 10% recomb.**
Ry (recombinant)	**ry = Rryy**	**12**	
rY (recombinant)	**ry = rrYy**	**8**	
ry (parental)	**ry = rryy**	**92**	**Map distance between loci R and Y = 10 cM**

Linkage mapping

EXAMPLE



Number of recombinants:
AC = 6/20 (1, 3, 10, 13, 18, 19)
AD = 4/20 (1, 3, 10, 13)
CD = 2/20 (18, 19)
C = 1/20 (plant number 18)
D = 1/20 (plant number 19)



Adapted from Paterson (1996). Explanations in next slide

Linkage mapping

EXAMPLE

Backcross:

Hybrid	**x**	**Recurrent parent**
AcD		**aCd**
aCd		**aCd**

Comparison of A and C:			
Gametes		***Progeny***	***Plant number***
Hybrid	Rec. parent		
Ac (P)	aC	Ac	
		aC	
aC (P)	aC	aC	
		aC	
AC (R)	aC	AC	1, 13, 18
		aC	
ac (R)	aC	ac	3, 10, 19
		aC	

Comparison of A and D:			
Gametes		***Progeny***	***Plant number***
Hybrid	Rec. parent		
AD (P)	ad	AD	
		ad	
ad (P)	ad	ad	
		ad	
Ad (R)	ad	Ad	1, 13
		ad	
aD (R)	ad	aD	3, 10
		ad	

Linkage mapping

DEFINITION

- Graphical representation of the genome of an organism
- A linear map of the relative positions of genes, molecular markers and phenotypic markers along a chromosome. Distances are established by linkage analysis, which determines the frequency at which two loci become separated during chromosomal recombination
- A genetic linkage map can be compared to a road map. Just as mile posts guide the motorist along a linear highway, molecular tools enable the geneticist to establish specific 'genetic markers' (DNA-markers) at defined places along each linear chromosome.

Linkage mapping

HISTORY

- Visual markers have been used in genetic studies since 1910
- Earliest linkage maps contained a few morphological markers mostly representing genes for which mutants were available
- At the DNA-marker level the amount of variation is so large that in a particular mapping population thousands of markers segregate
- Linkage maps were being used before it was known that DNA was the hereditary material, but DNA-based technology revolutionized genetic mapping in plants. This is because large numbers of genetic markers can be found by studying differences in the DNA molecule itself.

Linkage mapping

GENETIC BASIS

- "Genetic linkage", or co-transmission from parent to progeny of genetic markers (or genes or markers and genes) which are close together on the same chromosome, provides a means for determining the order of DNA markers along the chromosome
- By using in the analysis hundreds to thousands of markers, we can build up a 'linkage map' describing the relationships among the markers
- At the end we have a schematic representation of each chromosome, with the relative position of the DNA-markers

  - Only DNA-markers which are polymorphic in the studied population can be mapped => usually 'unrelated' genotypes are used as parents to construct the segregating population, to increase the probability that they will carry different alleles at many marker loci

Linkage mapping

TOOLS REQUIRED

- A mapping population. Different kinds of segregating populations can be used to construct a linkage map
- Hundreds to thousands of DNA-markers
- Appropriate statistical tools and software

Linkage mapping

STEPS

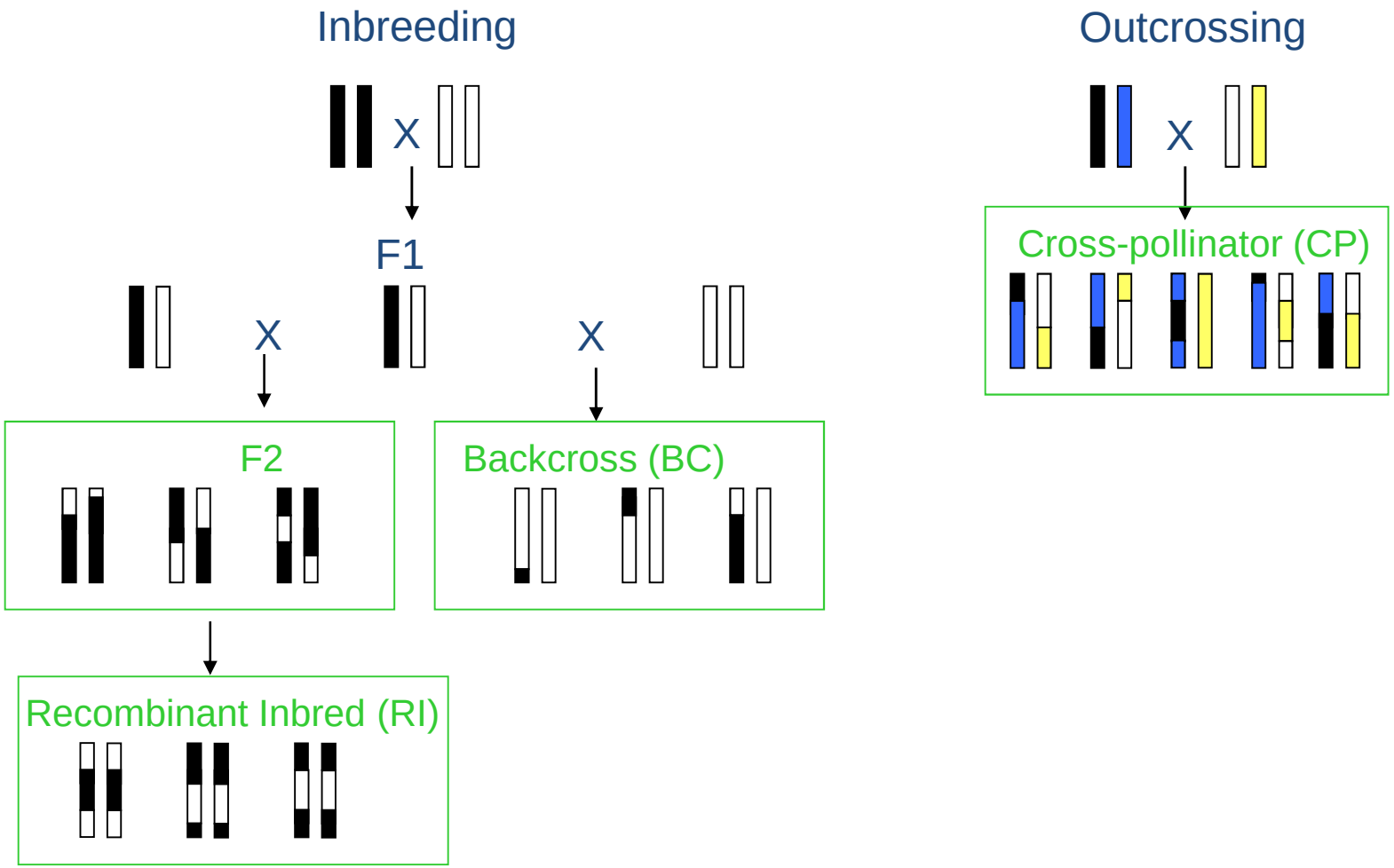
1. Creation of mapping population
2. Detection of polymorphic loci
3. Segregation analysis of single markers
4. Estimation of recombination frequencies
5. Identification of linkage groups
6. Calculation of map distances

Steps in linkage map construction

1. **Creation of mapping population**
2. Detection of polymorphic loci
3. Segregation analysis of single markers
4. Estimation of recombination frequencies
5. Identification of linkage groups
6. Calculation of map distances

Steps in linkage map construction

FEATURES OF DIFFERENT MAPPING POPULATIONS



Steps in linkage map construction

MAPPING POPULATIONS

- Backcross population: constructed by crossing the F1 hybrid to one of the parents (the “recurrent” parent). Only alleles derived from the “donor” (non-recurrent parent) segregate
- F2 population: can be constructed by selfing the F1 hybrid
- Double haploid (DH) population: made by regenerating plants from single (haploid) pollen grains produced by the F1, and inducing chromosome doubling
- Recombinant Inbred (RI) population: the progeny of an F2 cross are self-pollinated during several generations, by applying ‘single-seed-descent’
- Near-Isogenic (NI) population: the F1 hybrid is backcrossed to one of the parents (the “recurrent” parent) for various generations

Steps in linkage map construction

FEATURES OF DIFFERENT MAPPING POPULATIONS

	# possible genotypes per locus (diploids)	# generations to make	Replication?
DH	2	2	yes
Backcross	2	2	clonal only
F2	3	2	clonal only
RI	2	6 to 8	yes
Cross-pollinator	4	1	clonal only

Steps in linkage map construction

1. Creation of mapping population
2. **Detection of polymorphic loci**
3. Segregation analysis of single markers
4. Estimation of recombination frequencies
5. Identification of linkage groups
6. Calculation of map distances

Steps in linkage map construction

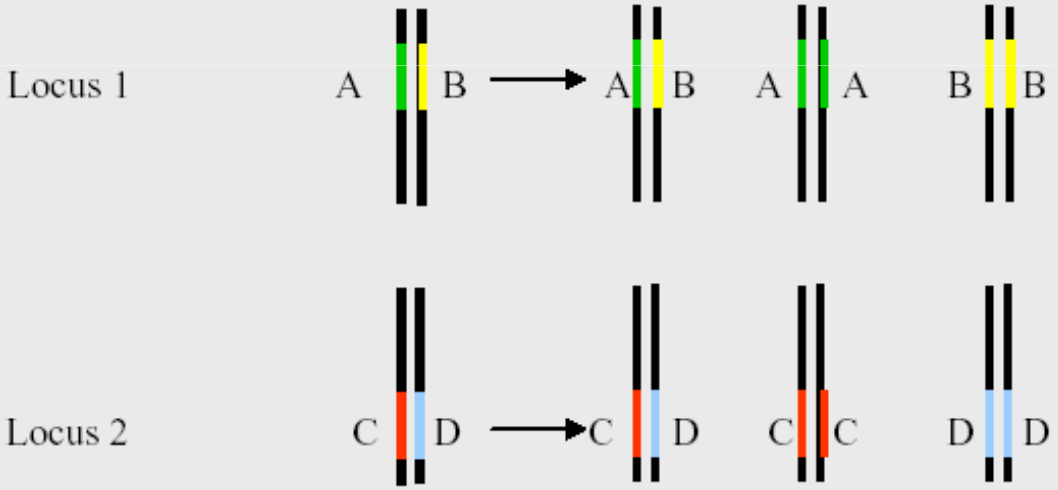
DETECTION OF POLYMORPHIC LOCI

RFLPs

SSR

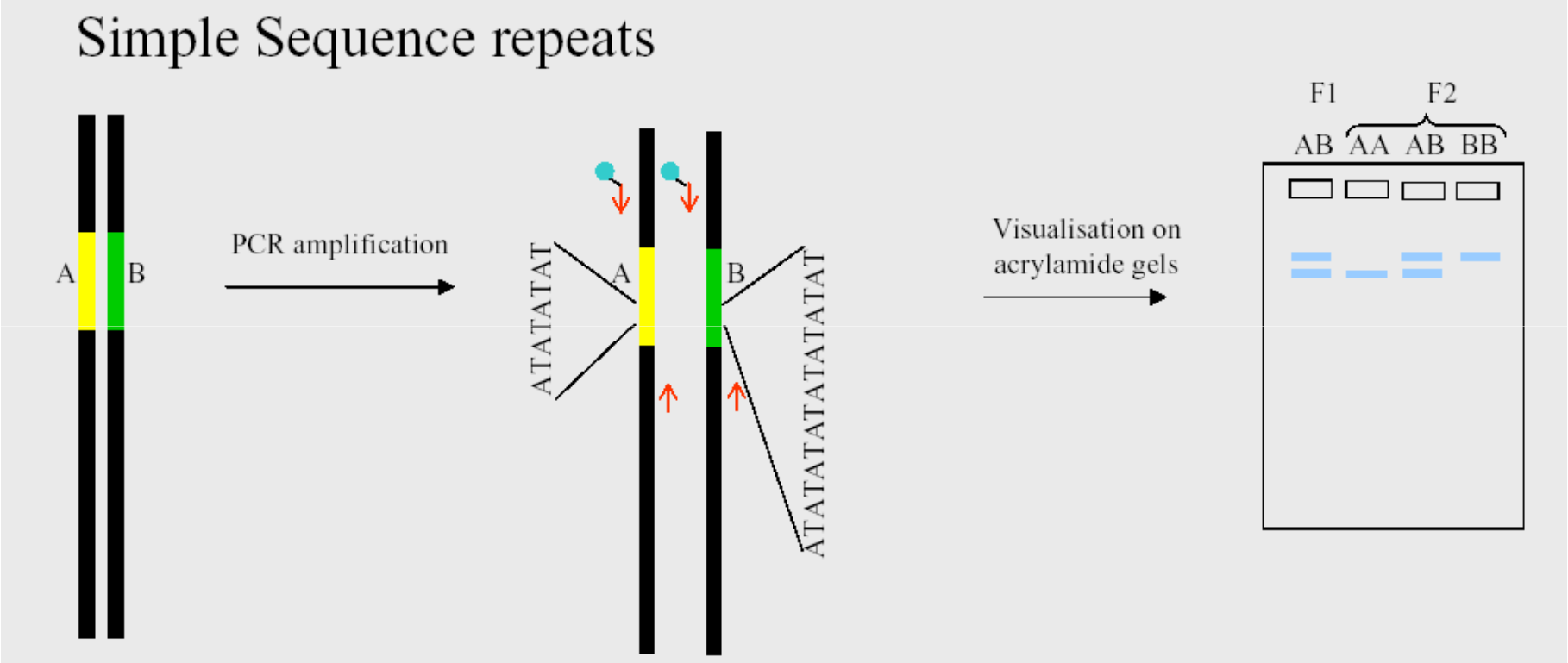
AFLP

STS



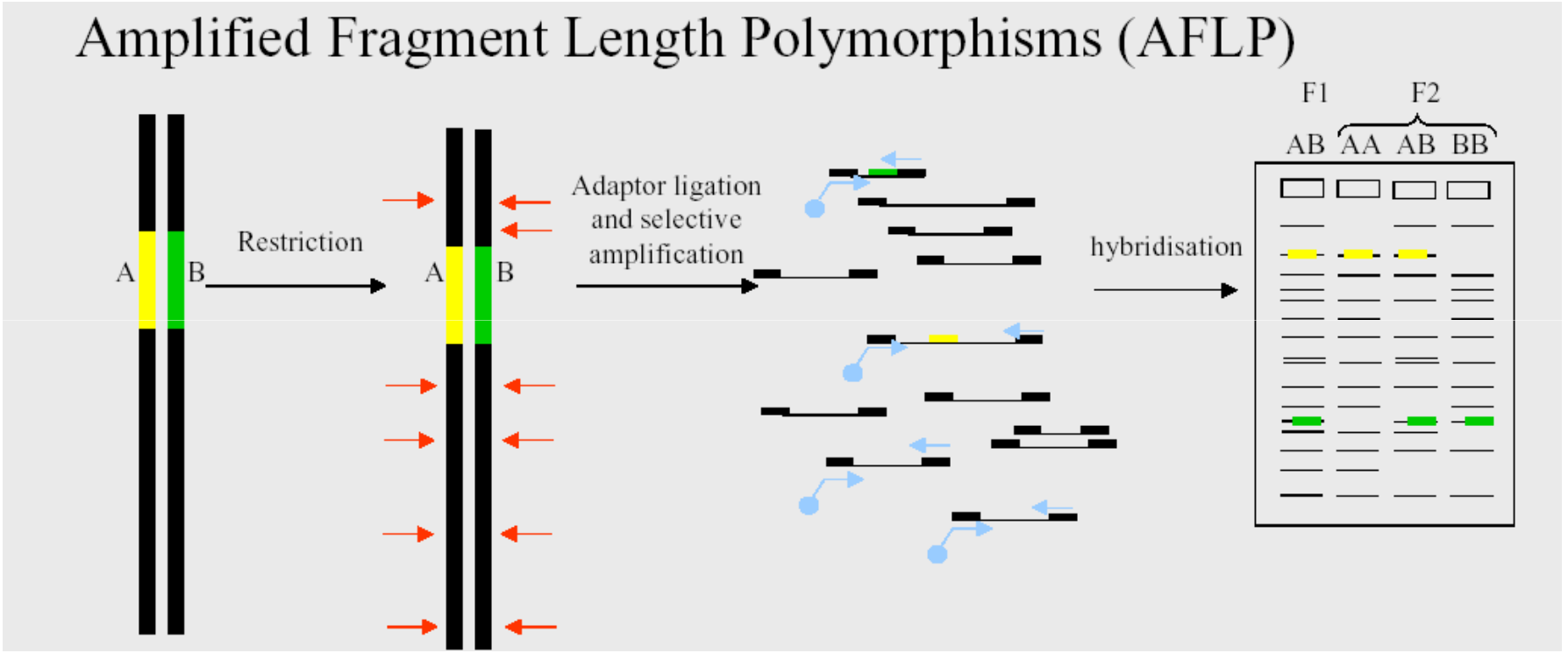
Steps in linkage map construction

DETECTION OF POLYMORPHIC LOCI



Steps in linkage map construction

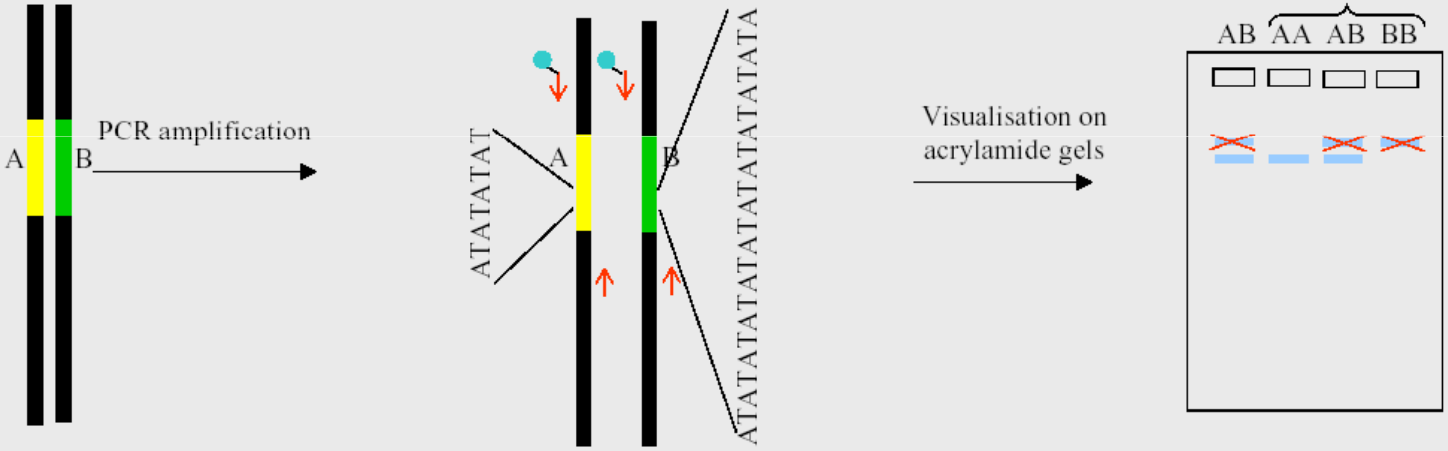
DETECTION OF POLYMORPHIC LOCI



Steps in linkage map construction

DETECTION OF POLYMORPHIC LOCI

- Dominant versus co-dominant markers
- Null-alleles can cause dominance



e.g. B allele is too long to amplify

⇒ No visible amplification product

⇒ No difference between AB and AA => A-

Steps in linkage map construction

1. Creation of mapping population
2. Detection of polymorphic loci
3. **Segregation analysis of single markers**
4. Estimation of recombination frequencies
5. Identification of linkage groups
6. Calculation of map distances

Steps in linkage map construction

SEGREGATION ANALYSIS OF SINGLE MARKERS

Expected genotype frequencies are defined per locus, according to the population structure and kind of marker (dominant/co-dominant)

Locus is AB in F1 => Expected segregation in F2 is 1:2:1

Locus is A- in F1 => Expected segregation in F2 is 3:1

Locus	Parent 1	Parent 2	F1	F2 individuals								
				1	2	3	4	5	6	7	8	9
1	AA	BB	AB	AA	AB	AB	AB	BB	BB	AA	AB	AB
2	AA	--	A-	--	A-	A-	--	--	A-	A-	A-	A-

Steps in linkage map construction

SEGREGATION ANALYSIS OF SINGLE MARKERS

- Check for deviation from expected segregation => X^2 test

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

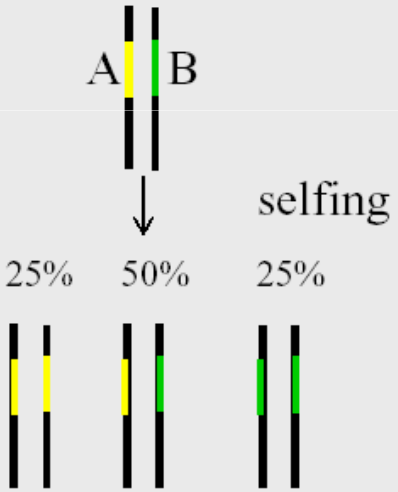
O = observed frequency

E = Expected frequency

Steps in linkage map construction

SEGREGATION ANALYSIS OF SINGLE MARKERS

- Example :



Genotypes	Expected	Observed	$(O-E)^2/E$
AA	25%	20%	1.00
AB	50%	40%	2.00
BB	25%	40%	9.00
		$\Sigma =$	12.00

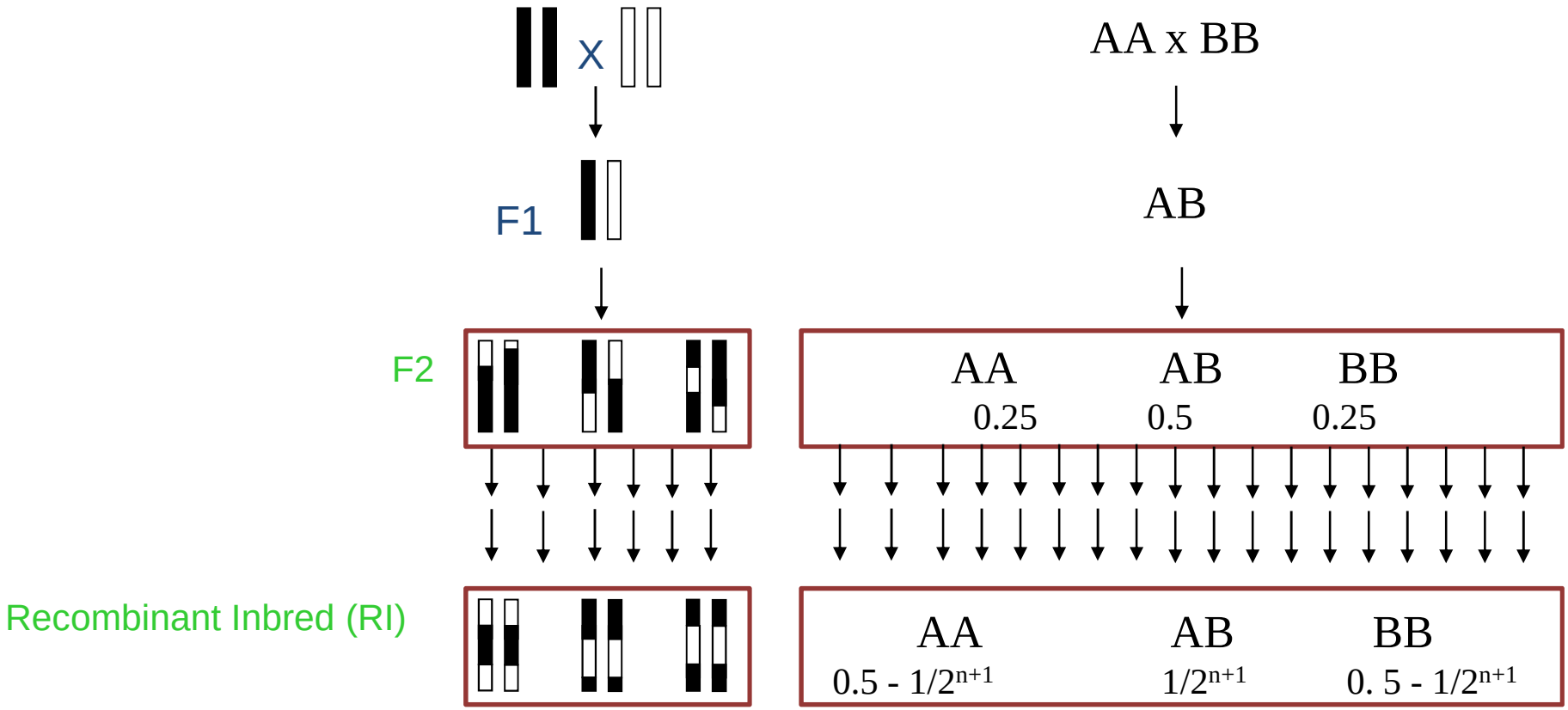
$\chi^2 = 12.00$ with 2 df

$\Rightarrow p<0.005$

$\Rightarrow$Hypothesis of 1:2:1 segregation is rejeceted

Example

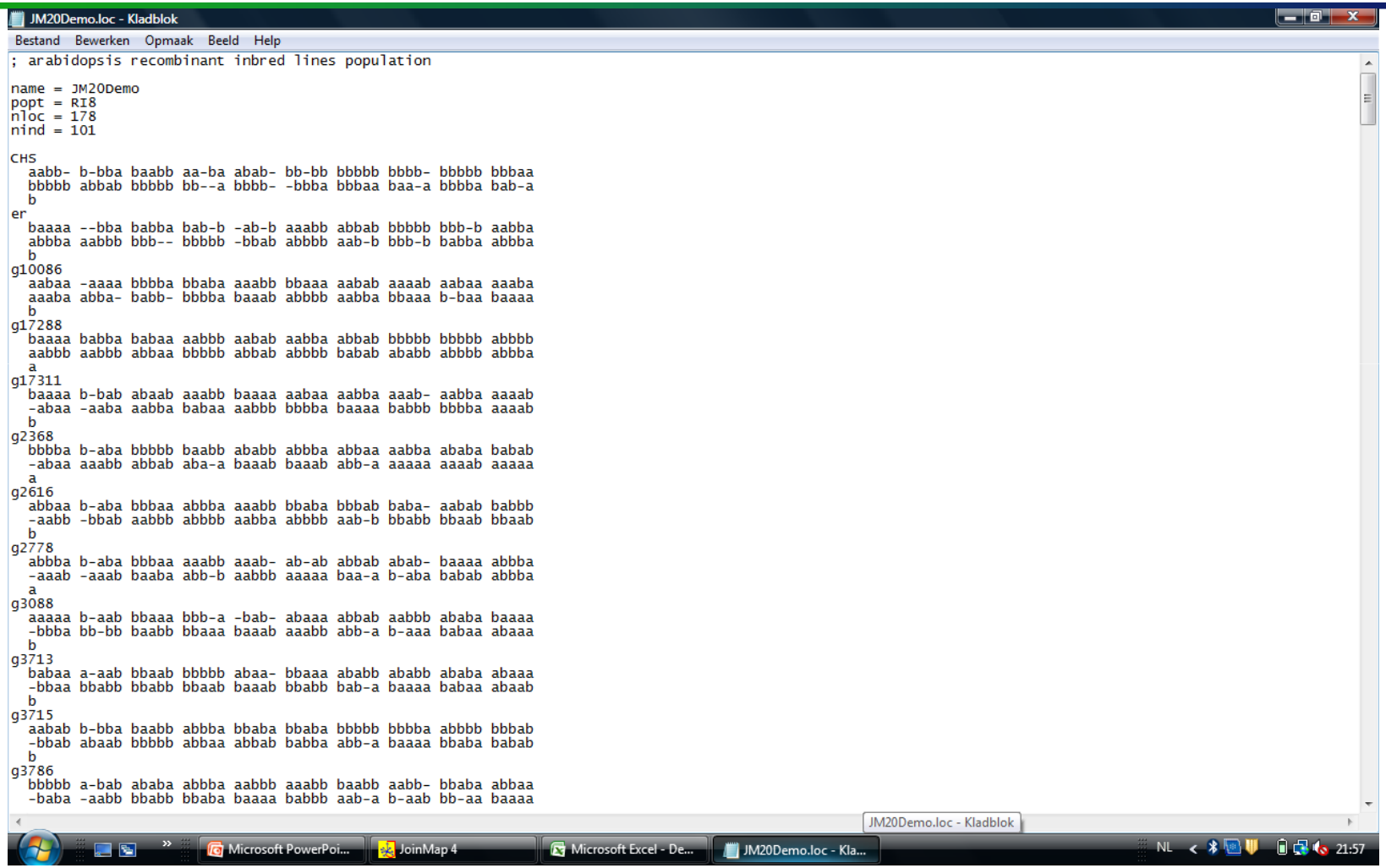
RIL



- e.g. Phaseolus vulgaris

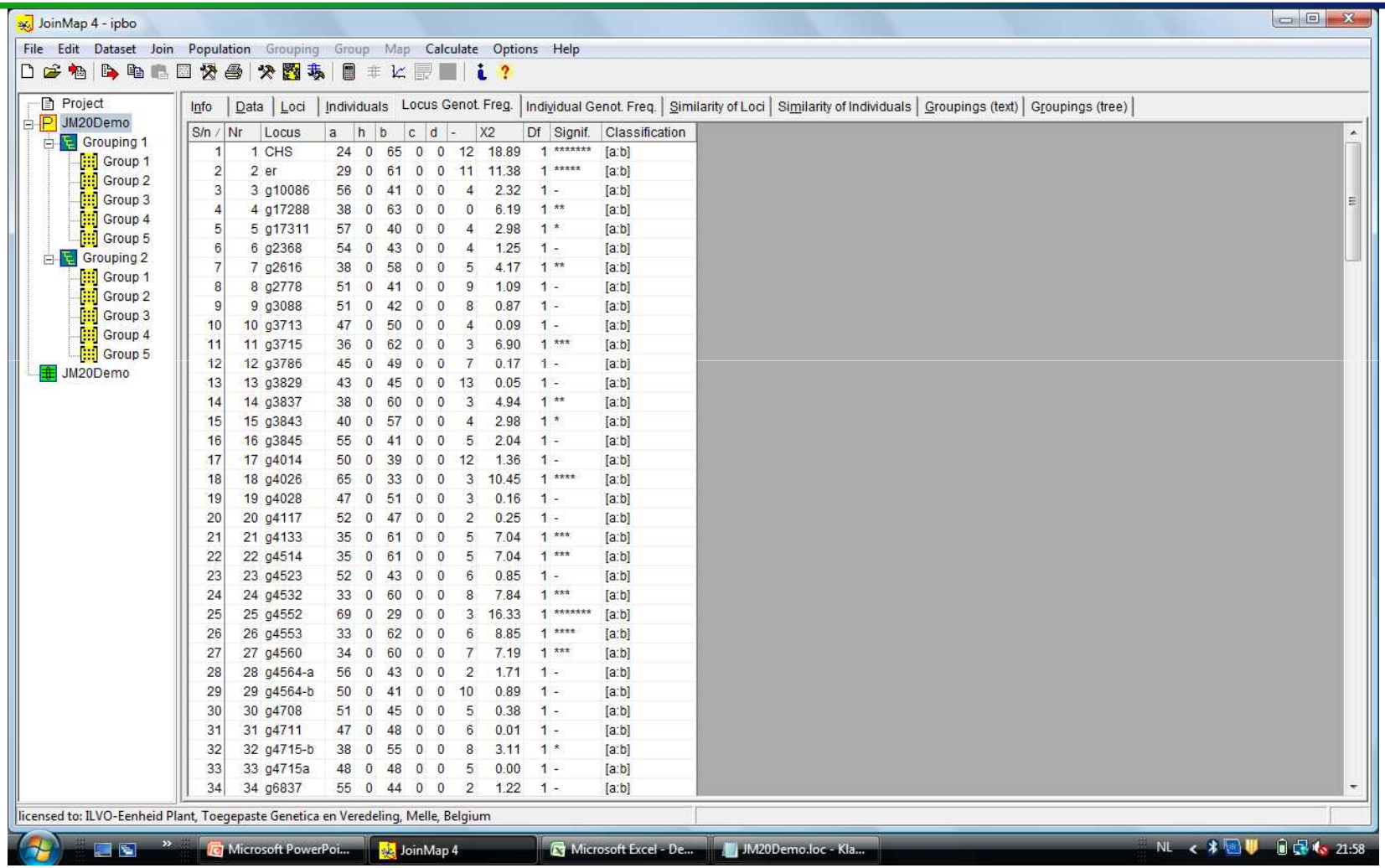
Example

RIL


JM20Demo.loc - Kladblok
Bestand Bewerken Opmaak Beeld Help
; arabidopsis recombinant inbred lines population
name = JM20Demo
popt = RI8
nloc = 178
nind = 101
CHS
aabb- b-bba baabb aa-ba abab- bb-bb bbbbb bbbb- bbbbb bbbaa
bbbbb abbab bbbbb bb--a bbbb- -bbba bbbaa baa-a bbbba bab-a
b
er
baaaa --bba babba bab-b -ab-b aaabb abbab bbbbb bbb-b aabba
abbba aabbb bbb-- bbbbb -bbab abbbb aab-b bbb-b babba abbba
b
g10086
aabaa -aaaa bbbba bbaba aaabb bbaaa aabab aaaab aabaa aaaba
aaaba abba- babb- bbbba baaab abbbb aabba bbaaa b-baa baaaa
b
g17288
baaaa babba babaa aabbb aabab aabba abbab bbbbb bbbbb abbbb
aabbb aabbb abbaa bbbbb abbab abbbb babab ababb abbbb abbba
a
g17311
baaaa b-bab abaab aaabb baaaa aabaa aabba aaab- aabba aaaab
-abaa -aaba aabba babaa aabbb bbbba baaaa babbb bbbba aaaab
b
g2368
bbbba b-aba bbbbb baabb ababb abbba abbaa aabba ababa babab
-abaa aaabb abbab aba-a baaab baaab abb-a aaaaa aaaab aaaaa
a
g2616
abbaa b-aba bbbaa abbba aaabb bbaba bbbab baba- aabab babbb
-aabb -bbab aabbb abbbb aabba abbbb aab-b bbabb bbaab bbaab
b
g2778
abbba b-aba bbbaa aaabb aaab- ab-ab abbab abab- baaaa abbba
-aaab -aaab baaba abb-b aabbb aaaaa baa-a b-aba babab abbba
a
g3088
aaaaa b-aab bbaaa bbb-a -bab- abaaa abbab aabbb ababa baaaa
-bbba bb-bb baabb bbaaa baaab aaabb abb-a b-aaa babaa abaaa
b
g3713
babaa a-aab bbaab bbbbb abaa- bbaaa ababb ababb ababa abaaa
-bbaa bbabb bbabb bbaab baaab bbabb bab-a baaaa babaa abaab
b
g3715
aabab b-bba baabb abbba bbaba bbaba bbbbb bbbba abbbb bbbab
-bbab abaab bbbbb abbaa abbab babba abb-a baaaa bbaba babab
b
g3786
bbbbb a-bab ababa abbba aabbb aaabb baabb aabb- bbaba abbaa
-baba -aabb bbabb bbaba baaaa babbb aab-a b-aab bb-aa baaaa
JM20Demo.loc - Kladblok
Microsoft PowerPoi...
JoinMap 4
Microsoft Excel - De...
JM20Demo.loc - Kla...
NL
21:57

Example

RIL



S/n /	Nr	Locus	a	h	b	c	d	-	X2	Df	Signif.	Classification
1	1	CHS	24	0	65	0	0	12	18.89	1	*******	[a:b]
2	2	er	29	0	61	0	0	11	11.38	1	*****	[a:b]
3	3	g10086	56	0	41	0	0	4	2.32	1	-	[a:b]
4	4	g17288	38	0	63	0	0	0	6.19	1	**	[a:b]
5	5	g17311	57	0	40	0	0	4	2.98	1	*	[a:b]
6	6	g2368	54	0	43	0	0	4	1.25	1	-	[a:b]
7	7	g2616	38	0	58	0	0	5	4.17	1	**	[a:b]
8	8	g2778	51	0	41	0	0	9	1.09	1	-	[a:b]
9	9	g3088	51	0	42	0	0	8	0.87	1	-	[a:b]
10	10	g3713	47	0	50	0	0	4	0.09	1	-	[a:b]
11	11	g3715	36	0	62	0	0	3	6.90	1	***	[a:b]
12	12	g3786	45	0	49	0	0	7	0.17	1	-	[a:b]
13	13	g3829	43	0	45	0	0	13	0.05	1	-	[a:b]
14	14	g3837	38	0	60	0	0	3	4.94	1	**	[a:b]
15	15	g3843	40	0	57	0	0	4	2.98	1	*	[a:b]
16	16	g3845	55	0	41	0	0	5	2.04	1	-	[a:b]
17	17	g4014	50	0	39	0	0	12	1.36	1	-	[a:b]
18	18	g4026	65	0	33	0	0	3	10.45	1	****	[a:b]
19	19	g4028	47	0	51	0	0	3	0.16	1	-	[a:b]
20	20	g4117	52	0	47	0	0	2	0.25	1	-	[a:b]
21	21	g4133	35	0	61	0	0	5	7.04	1	***	[a:b]
22	22	g4514	35	0	61	0	0	5	7.04	1	***	[a:b]
23	23	g4523	52	0	43	0	0	6	0.85	1	-	[a:b]
24	24	g4532	33	0	60	0	0	8	7.84	1	***	[a:b]
25	25	g4552	69	0	29	0	0	3	16.33	1	*******	[a:b]
26	26	g4553	33	0	62	0	0	6	8.85	1	****	[a:b]
27	27	g4560	34	0	60	0	0	7	7.19	1	***	[a:b]
28	28	g4564-a	56	0	43	0	0	2	1.71	1	-	[a:b]
29	29	g4564-b	50	0	41	0	0	10	0.89	1	-	[a:b]
30	30	g4708	51	0	45	0	0	5	0.38	1	-	[a:b]
31	31	g4711	47	0	48	0	0	6	0.01	1	-	[a:b]
32	32	g4715-b	38	0	55	0	0	8	3.11	1	*	[a:b]
33	33	g4715a	48	0	48	0	0	5	0.00	1	-	[a:b]
34	34	g6837	55	0	44	0	0	2	1.22	1	-	[a:b]

Steps in linkage map construction

SEGREGATION ANALYSIS OF SINGLE MARKERS

- Deviation from Mendelian segregation is called segregation distortion
- Causes:
  - selection
  - competition of gametes
  - non-random mating
  - ….
- Consequences:
  - innacurate estimation of recombination frequencies between markers
  - Innacurate estimation of QTL positions and effects
  - …

Usually only markers that fit expected genotype proportions are kept for further steps

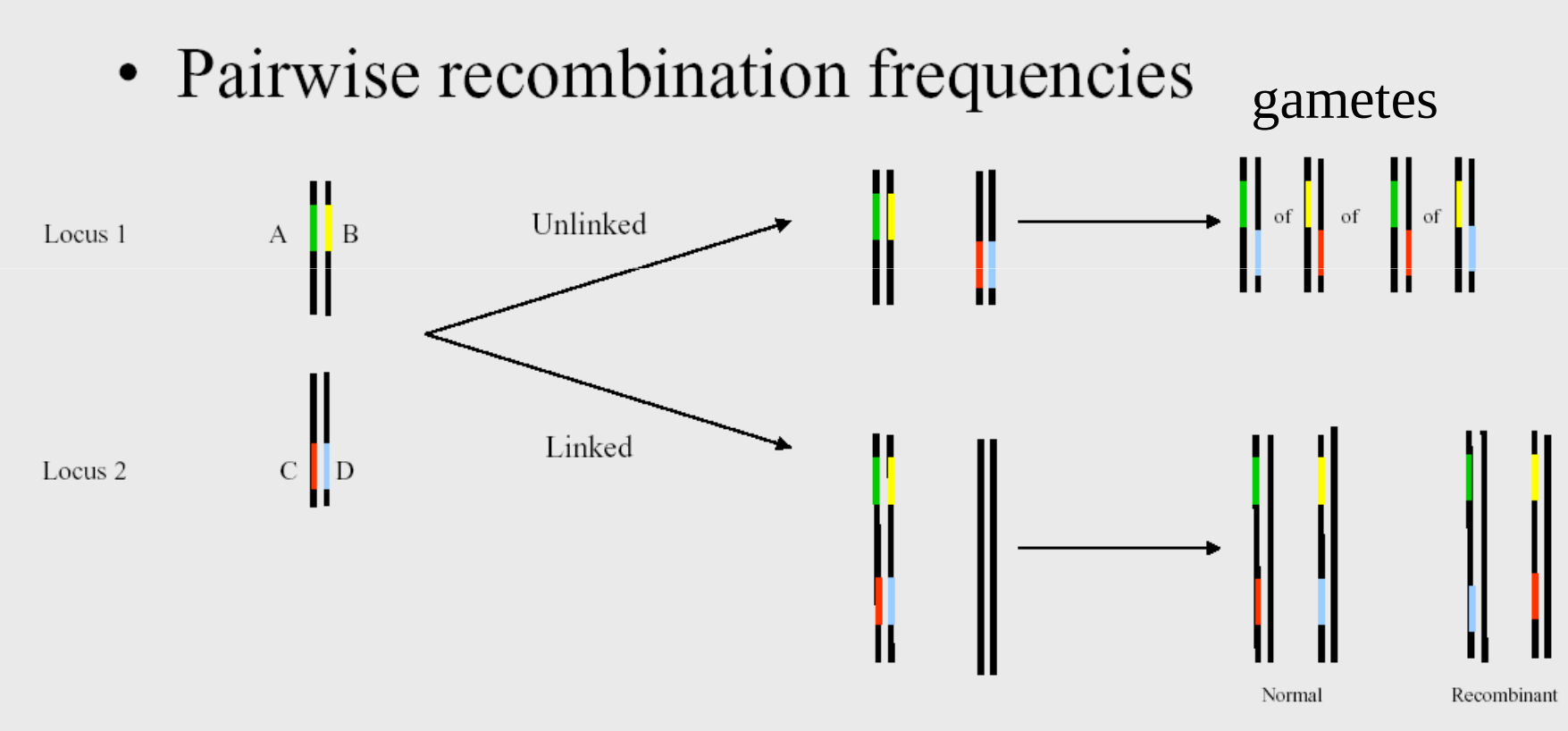
Steps in linkage map construction

1. Creation of mapping population
2. Detection of polymorphic loci
3. Segregation analysis of single markers
4. **Estimation of recombination frequencies**
5. Identification of linkage groups
6. Calculation of map distances

Steps in linkage map construction

ESTIMATION OF RECOMBINATION FREQUENCIES

A diploid plant with two chromosomes (2n=2x=4)



Steps in linkage map construction

ESTIMATION OF RECOMBINATION FREQUENCIES

Parents: AB/AB (donor) x ab/ab (recurrent)

F1: AB/ab

gametes	AB	Ab	aB	ab
ab	ab/AB	ab/Ab	ab/aB	ab/ab

BC1 genotypes	ab/AB	ab/Ab	ab/aB	ab/ab
frequencies	(1-r)/2	r/2	r/2	(1-r)/2

Steps in linkage map construction

ESTIMATION OF RECOMBINATION FREQUENCIES

Estimation in a BC population:

gametes	AB	Ab	aB	ab
ab	ab/AB	ab/Ab	ab/aB	ab/ab

Genotype	Observed numbers	Expected freq
ab/AB	N1	(1-r)/2
ab/Ab	N2	(r/2)
ab/aB	N3	(r/2)
ab/ab	N4	(1-r)/2

$$r = (N2+N3)/(N1+N2+N3+N4)$$

Steps in linkage map construction

ESTIMATION OF RECOMBINATION FREQUENCIES

- Loci located in close (physical) proximity => small chance of recombination

- The number of crossovers between two loci

  => estimate the map distance (centiMorgans)

- Recombination events can be recognized only the basis of haplotypes

  => determine haplotypes in mapping populations

Steps in linkage map construction

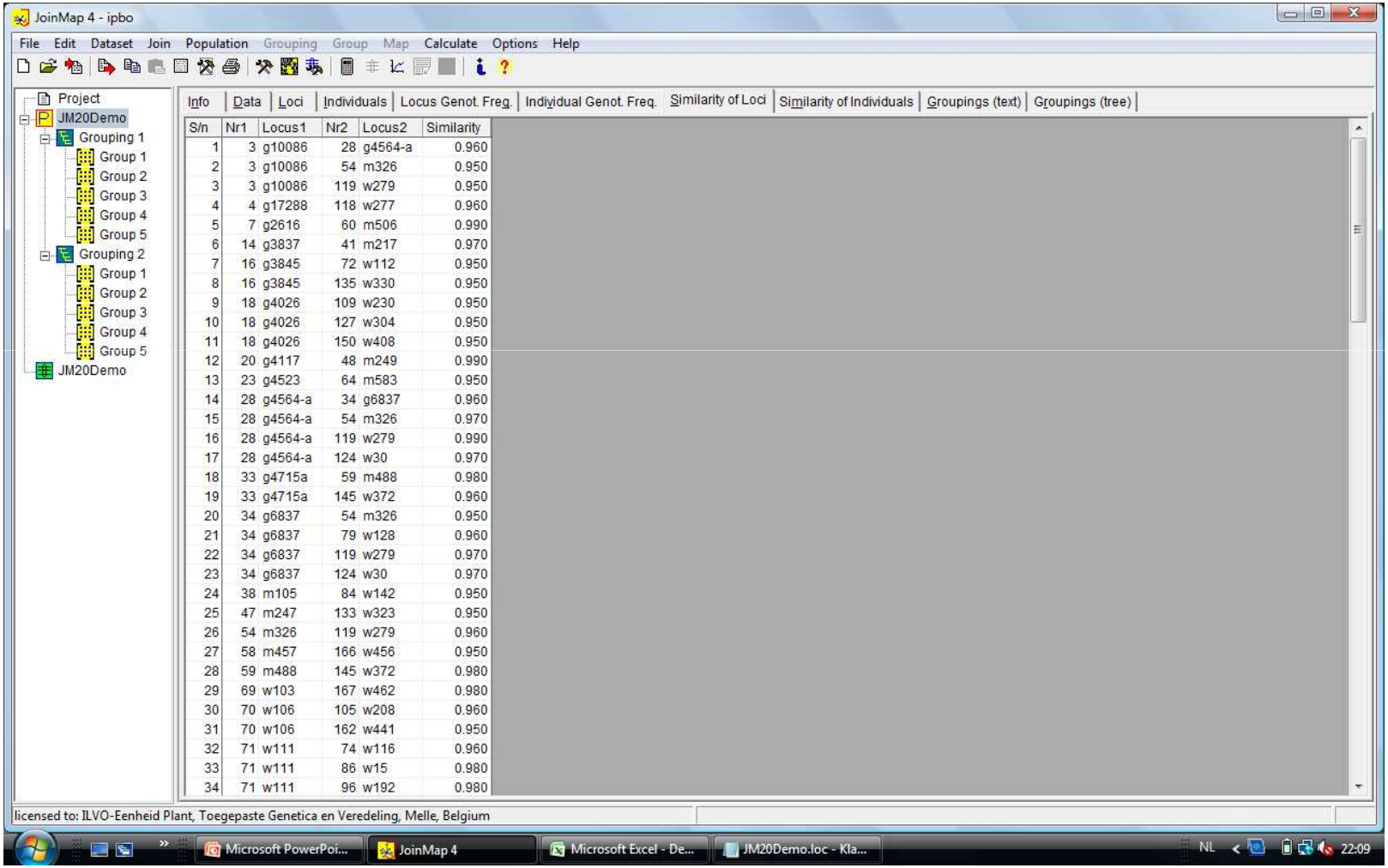
ESTIMATION OF RECOMBINATION FREQUENCIES

LOD-Score

- Maximum likelihood (ML)
  - Generic name of statistical approaches in which one aims to find the parameter value (the value of *r* in our case) that maximizes the likelihood of the data. The question is: given the observed genotype values, what can we say about r?
  - $L(r \mid x)$= likelihood of r given the observed vector of data x
  - The MLE (maximum likelihood estimate) is the value of r that is most likely to have produced the data observed
- Likelihood Ratio test (LRT)
  - The ratio between the likelihood of r taking value r_1 (MLE), and the likelihood of r under the null hypothesis r_0 (r_0 =0.5; no linkage)
  - $L1 / L0 = L(r_1 \mid x) / L(r_0 \mid x)$
- LOD score (Z)
  - $Z = \log_{10} (L(r_1 \mid x) / L(r_0 \mid x))$
  - The alternative hypothesis is 10^Z times more likely than the null hypothesis (Z = 3 means that the alternative hypothesis is 10^3 times more likely than the null hypothesis)
  - In general, when Z = 3 or higher linkage is considered significant

Example

RIL



S/n	Nr1	Locus1	Nr2	Locus2	Similarity
1	3	g10086	28	g4564-a	0.960
2	3	g10086	54	m326	0.950
3	3	g10086	119	w279	0.950
4	4	g17288	118	w277	0.960
5	7	g2616	60	m506	0.990
6	14	g3837	41	m217	0.970
7	16	g3845	72	w112	0.950
8	16	g3845	135	w330	0.950
9	18	g4026	109	w230	0.950
10	18	g4026	127	w304	0.950
11	18	g4026	150	w408	0.950
12	20	g4117	48	m249	0.990
13	23	g4523	64	m583	0.950
14	28	g4564-a	34	g6837	0.960
15	28	g4564-a	54	m326	0.970
16	28	g4564-a	119	w279	0.990
17	28	g4564-a	124	w30	0.970
18	33	g4715a	59	m488	0.980
19	33	g4715a	145	w372	0.960
20	34	g6837	54	m326	0.950
21	34	g6837	79	w128	0.960
22	34	g6837	119	w279	0.970
23	34	g6837	124	w30	0.970
24	38	m105	84	w142	0.950
25	47	m247	133	w323	0.950
26	54	m326	119	w279	0.960
27	58	m457	166	w456	0.950
28	59	m488	145	w372	0.980
29	69	w103	167	w462	0.980
30	70	w106	105	w208	0.960
31	70	w106	162	w441	0.950
32	71	w111	74	w116	0.960
33	71	w111	86	w15	0.980
34	71	w111	96	w192	0.980

Steps in linkage map construction

1. Creation of mapping population
2. Detection of polymorphic loci
3. Segregation analysis of single markers
4. Estimation of recombination frequencies
5. **Identification of linkage groups**
6. Calculation of map distances

Steps in linkage map construction

IDENTIFICATION OF LINKAGE GROUPS

- Use pairwise recombination frequency of all marker pairs

- Define groups of markers with a high likelihood to segregate together

  => at increasing stringency levels of a test for linkage

  => using a LOD score as threshold

Steps in linkage map construction

IDENTIFICATION OF LINKAGE GROUPS

- Different LOD thresholds => different numbers of linkage groups

  => LOD score low => low number of linkage groups

  => LOD score high => high number of linkage groups

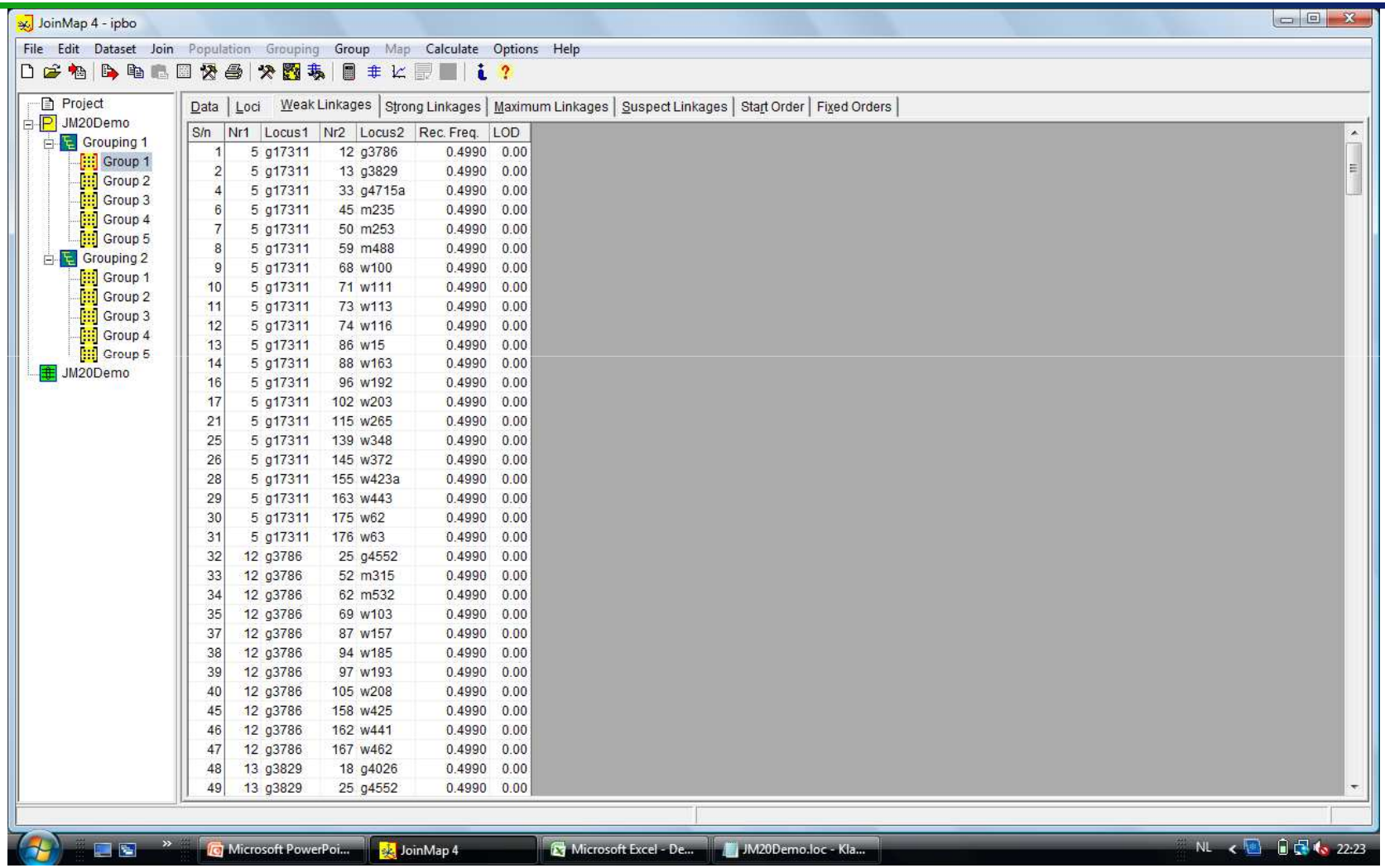
- LOD values between 4 and 7 usually the ‘best’ grouping
- Ideally the number of linkage groups equals the haploid chromosome number over a wide range of LOD values

  => BUT if too few markers

  => more linkage groups than haploid chromosome number

Example

RIL


JoinMap 4 - ipbo
File Edit Dataset Join Population Grouping Group Map Calculate Options Help
Project
JM20Demo
Grouping 1
Group 1
Group 2
Group 3
Group 4
Group 5
Grouping 2
Group 1
Group 2
Group 3
Group 4
Group 5
JM20Demo
Data Loci Weak Linkages Strong Linkages Maximum Linkages Suspect Linkages Start Order Fixed Orders
S/n Nr1 Locus1 Nr2 Locus2 Rec. Freq. LOD
1 5 g17311 12 g3786 0.4990 0.00
2 5 g17311 13 g3829 0.4990 0.00
4 5 g17311 33 g4715a 0.4990 0.00
6 5 g17311 45 m235 0.4990 0.00
7 5 g17311 50 m253 0.4990 0.00
8 5 g17311 59 m488 0.4990 0.00
9 5 g17311 68 w100 0.4990 0.00
10 5 g17311 71 w111 0.4990 0.00
11 5 g17311 73 w113 0.4990 0.00
12 5 g17311 74 w116 0.4990 0.00
13 5 g17311 86 w15 0.4990 0.00
14 5 g17311 88 w163 0.4990 0.00
16 5 g17311 96 w192 0.4990 0.00
17 5 g17311 102 w203 0.4990 0.00
21 5 g17311 115 w265 0.4990 0.00
25 5 g17311 139 w348 0.4990 0.00
26 5 g17311 145 w372 0.4990 0.00
28 5 g17311 155 w423a 0.4990 0.00
29 5 g17311 163 w443 0.4990 0.00
30 5 g17311 175 w62 0.4990 0.00
31 5 g17311 176 w63 0.4990 0.00
32 12 g3786 25 g4552 0.4990 0.00
33 12 g3786 52 m315 0.4990 0.00
34 12 g3786 62 m532 0.4990 0.00
35 12 g3786 69 w103 0.4990 0.00
37 12 g3786 87 w157 0.4990 0.00
38 12 g3786 94 w185 0.4990 0.00
39 12 g3786 97 w193 0.4990 0.00
40 12 g3786 105 w208 0.4990 0.00
45 12 g3786 158 w425 0.4990 0.00
46 12 g3786 162 w441 0.4990 0.00
47 12 g3786 167 w462 0.4990 0.00
48 13 g3829 18 g4026 0.4990 0.00
49 13 g3829 25 g4552 0.4990 0.00
Microsoft PowerPoi...
JoinMap 4
Microsoft Excel - De...
JM20Demo.loc - Kla...
NL
22:23

Steps in linkage map construction

1. Creation of mapping population
2. Detection of polymorphic loci
3. Segregation analysis of single markers
4. Estimation of recombination frequencies
5. Identification of linkage groups
6. **Calculation of map distances**

Steps in linkage map construction

CALCULATE MAP DISTANCES

- Once we know which markers belong to the same linkage group and which is the pair-wise recombination between all of them, we can order the markers along the linkage group
- To do this, we try to find the marker order that minimizes the number of crossovers
- Several approaches available in mapping programs

rAB = 0.10
rBC = 0.07
rAC = 0.15

Steps in linkage map construction

CALCULATE MAP DISTANCES

- The distance along a genetic map is derived from the frequency of recombination between genetic markers. 1 cM (*centiMorgan*) corresponds to an average of 1 recombinant in 100 gametes (1% recombination)
  - Usually, the distance is adjusted for the possibility of 'double recombination', which makes individuals to appear to represent a parental type, but in fact containing two recombination events
  - The probability of double recombination is proportional to the square of the recombination distance between two loci.
- The precision with which genetic distance is measured, is directly related to the number of individuals which is studied (if no recombinants found a sample of 20 progeny plants => recombination fraction = 0; but analyzing 80 additional individuals 1 recombinant can appear => recombination fraction = 1).
  - Typically a primary genetic map is constructed based on 50-100 individuals, permitting to detect recombination between markers 1-3 cM appart.
  - If higher precision is required, more individuals should be analyzed

Steps in linkage map construction

CALCULATE MAP DISTANCES

- Map function translates recombination frequencies into map distances
  - Haldane function

$$X = \tfrac{1}{2} \ln (1-2r)$$

- Takes into account double strand , three and four strand crossovers
- BUT assumes that 2-3-4 strand crossovers does not interfere with single crossovers

Steps in linkage map construction

CALCULATE MAP DISTANCES

- Map function translates recombination frequencies into map distances
  - Kosambi function
    - Takes into account interference

$$X = \tfrac{1}{4} \ln [(1+2r)/(1-2r)]$$

Steps in linkage map construction

CALCULATE MAP DISTANCES

- The genetic distance is only loosely related to the physical distance

  => influenced by - genetic
  - epigenetic
  - environmental factors

- For example, repetitive DNA elements are relatively inert in recombination, recombination is also reduced around the centromers

Steps in linkage map construction

CALCULATE MARKER ORDER

- There are many possibilities, but one of the computationally ‘cheapest’ algorithms is the greedy algorithm (stepwise build-up of the map by adding one marker at a time):
  - Start with two markers A and B
  - Add third marker C at the three different positions which are possible: CAB, ACB, ABC
  - The three orders are compared and the best fitting one is chosen (ACB)
  - Then add a fourth marker D at the four different positions which are possible: DACB, ADCB, ACDB, ACBD
  - The four orders are compared and the best fitting one is chosen

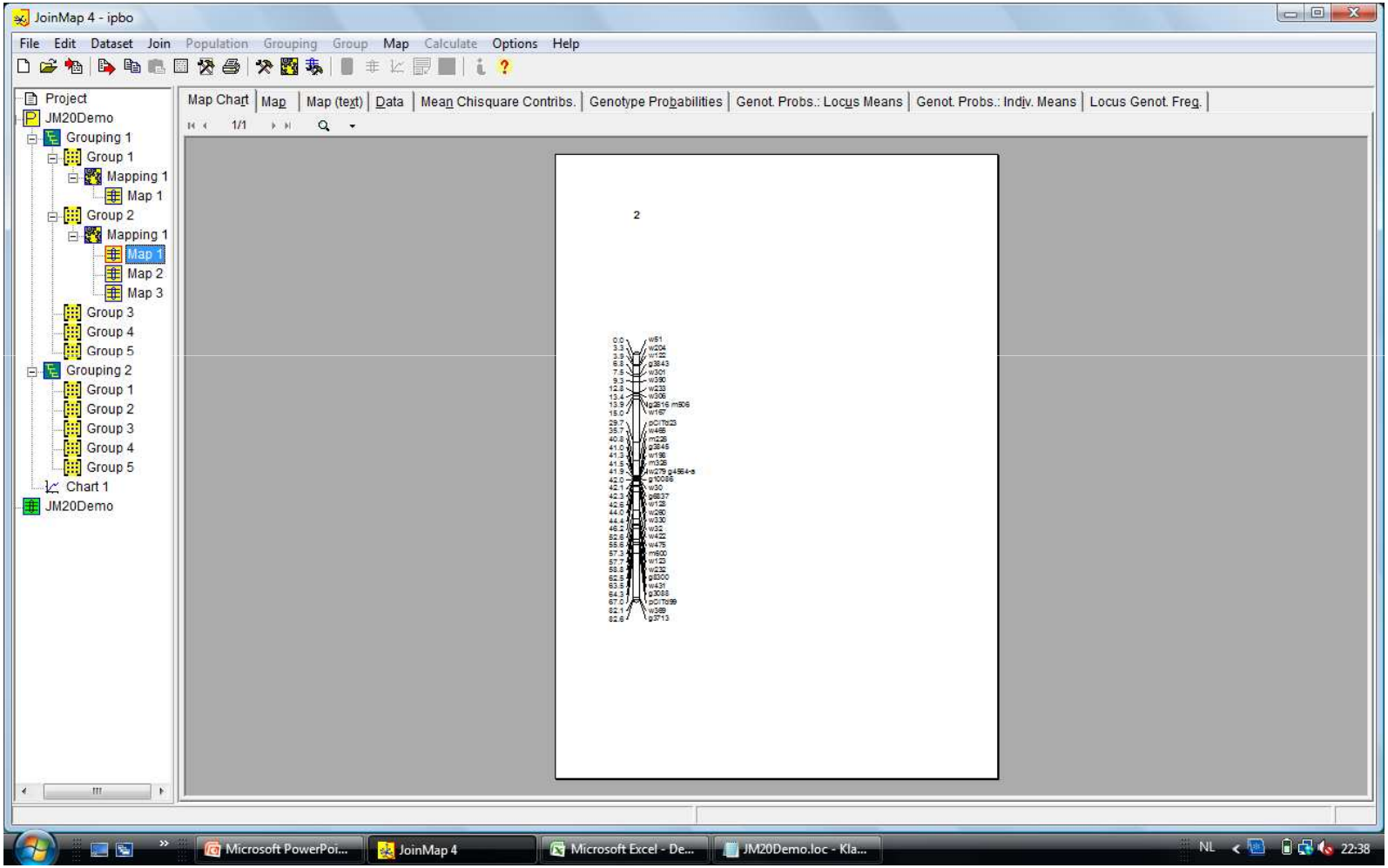
- JoinMap uses this algorithm, but with several refinements:
  - The order in which the markers are added is not random
  - After a marker has been added a ‘local reshuffling’ is applied to prevent that the previous sequence will not be changed any more

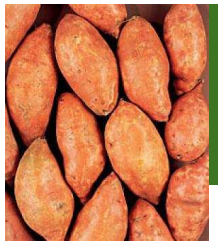
- There are many possibilities to use ad ‘fitting criterion’ (minimization of the sum of adjacent distances, maximization of the sum of adjacent LOD scores, weighted least squares)

ILVO

Example

RIL

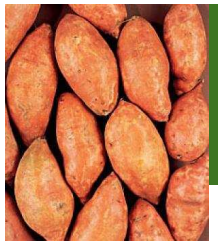

JoinMap 4 - ipbo
File Edit Dataset Join Population Grouping Group Map Calculate Options Help
Project
JM20Demo
Grouping 1
Group 1
Mapping 1
Map 1
Group 2
Mapping 1
Map 1
Map 2
Map 3
Group 3
Group 4
Group 5
Grouping 2
Group 1
Group 2
Group 3
Group 4
Group 5
Chart 1
JM20Demo
Map Chart | Map | Map (text) | Data | Mean Chisquare Contribs. | Genotype Probabilities | Genot. Probs.: Locus Means | Genot. Probs.: Indiv. Means | Locus Genot. Freq.
1/1
2
Microsoft PowerPoi...
JoinMap 4
Microsoft Excel - De...
JM20Demo.loc - Kla...
NL
22:38



Mapping in autopolyploids

MAPPING POPULATIONS IN POLYPLOIDS

	# possible genotypes per locus (diploids)	# generations to make	Replication?
DH	2	2	yes
Backcross	2	2	clonal only
F2	3	2	clonal only
RI	2	6 to 8	yes
Cross-pollinator	4	1	clonal only
Sweetpotato	400!!!!!	1	clonal only



Mapping in autopolyploids

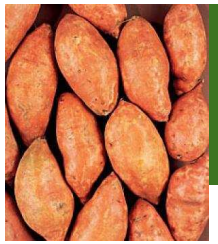
MAPPING POPULATIONS IN POLYPLOIDS

Why is mapping in polyploids more complicated than in diploids?

1. Lage number of genotypes expected (up to 400 if 12 alleles segregating)

2. Genotype of an individual is not always readily inferred (even using co-dominant markers)
   - Result of genotyping at one locus = ABC
   - AAAABC, ABBBCC, AABBCC, …..?

3. The type of ploidy (allopolyploidy or autopolyploidy) of many crops is unclear
   - Wheat is an allopolyploid (mapping is as in a diploid with many chromosomes)
   - Sweetpotato is an autopolyploid with hexasomic inheritance (Cervantes-Flores et al 2008)



Mapping in autopolyploids

MAPPING POPULATIONS IN POLYPLOIDS

Chromosome segregation in polyploids:

- Disomic (wheat)
- Hexasomic (sweetpotato)

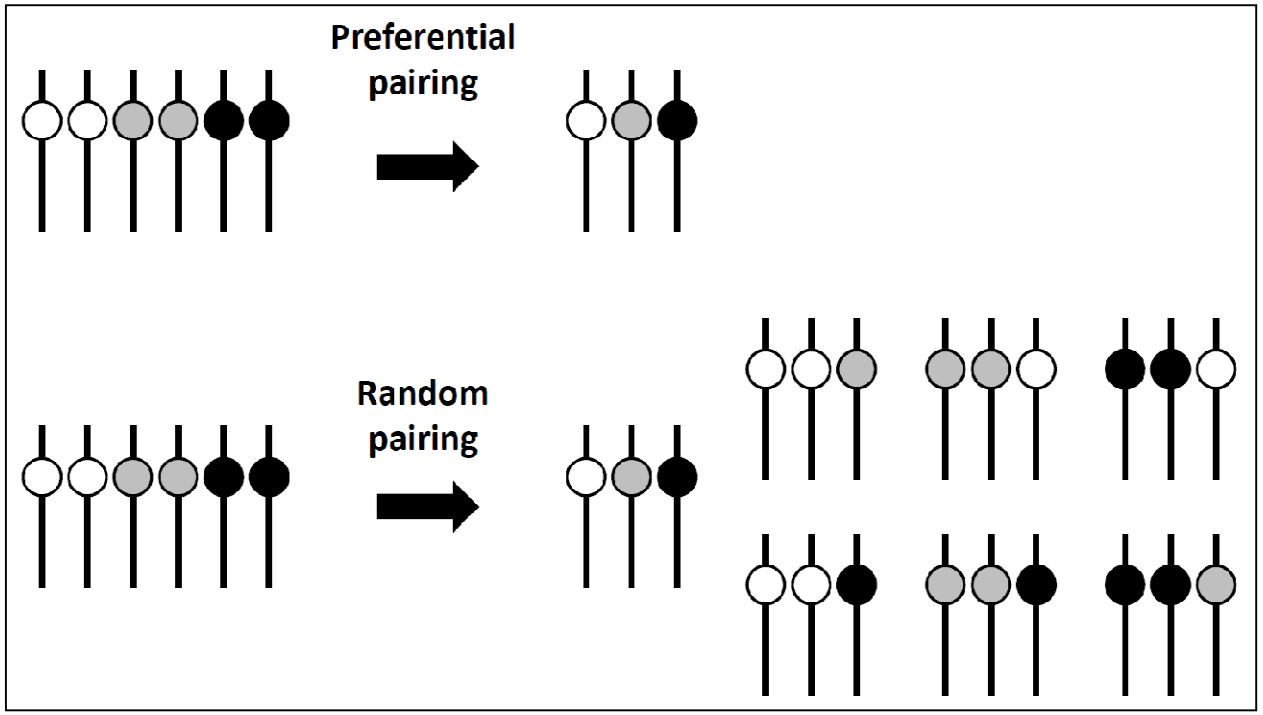
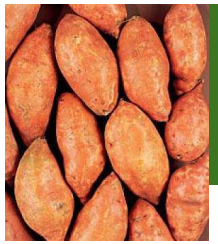


Figure: De Baker (2012)



Mapping in autopolyploids

SEGREGATION ANALYSIS OF SINGLE MARKERS

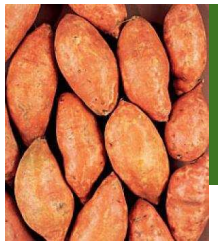
Kreigner et al (2003) - hexasomic inheritance in sweetpotato

Table 1. Expected segregation ratios (presence:absence) for the inheritance of a dominant marker in hexaploid sweetpotato, according to four cytological hypotheses (Jones 1967)

Marker dose	**Hypothesis I** Autohexaploid (hexasomic)		**Hypothesis II and III** Tetradiploid (tetra-disomic, tetra-somic, disomic)		**Hypothesis IV** Allohexaploid (disomic)	
Simplex	Aaaaaa	1:1	Aaaa aa	1:1	Aa aa aa	1:1
			aaaa Aa	1:1		
Duplex	AAaaaa	4:1	AAaa aa	5:1 [2)]	Aa Aa aa	3:1
			Aaaa Aa	3:1 [3)]	AA aa aa	1:0
			aaaa AA	1:0 [1)]		
Triplex	AAAaaa	19:1	AAAa aa	–	Aa Aa Aa	7:1
			AAaa Aa	11:1	AA Aa aa	1:0
			Aaaa AA	1:0		
Quadruplex	AAAAaa	1:0	AAAA aa	1:0	AA Aa Aa	1:0

[1)] disomic inheritance
[2)] tetrasomic inheritance
[3)] tetra-disomic inheritance

Using this table, it is possible to check the segregation of individual dominant markers in sweetpotato

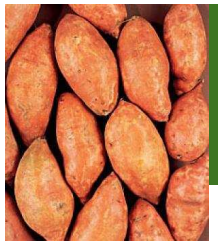


Mapping in autopolyploids

SEGREGATION ANALYSIS OF SINGLE MARKERS

- Deviation from Mendelian segregation is called segregation distortion
- Causes:
  - selection
  - competition of gametes
  - non-random mating
  - non-random pairing of chromosomes (in autopolyploids)
  - ….
- Consequences:
  - innacurate estimation of recombination frequencies between markers
  - Innacurate estimation of QTL positions and effects
  - …

ILVO



Mapping in autopolyploids

ESTIMATION OF RECOMBINATION FREQUENCIES

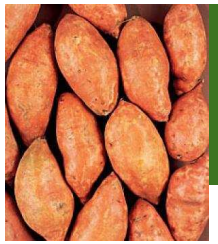
For tetraploids, suited linkage mapping program available: Tetraploidmap (http://www.bioss.ac.uk/knowledge/tetraploidmap/)

- Handles dominant and co-dominant markers in all possible configurations, taking account of null alleles
- Free to download

Approach in hexaploids (Cervantes-Flores et al 2008):

- Double pseudotestcross strategy
    - Reconstruct individual chrosomosomes in each parent independently
- Simplex AFLP markers (=SDF) are present in one parent in a single copy => segregate 50%/50% (present:absent) in progeny
    - A00000 x 000000 => A000000 / 0000000
- Mapping in this case is similar to diploid situation and can be carried out using standard mapping programs such as JoinMap or MapMaker
- Limitation: using simplex markers, impossible to identify homologous chromosomes

ILVO



Mapping in autopolyploids

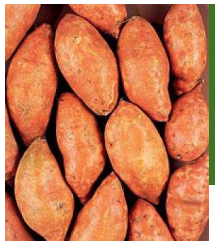
ESTIMATION OF RECOMBINATION FREQUENCIES

Simplex markers in coupling allow to reconstruct individual chromosomes in each parent separately

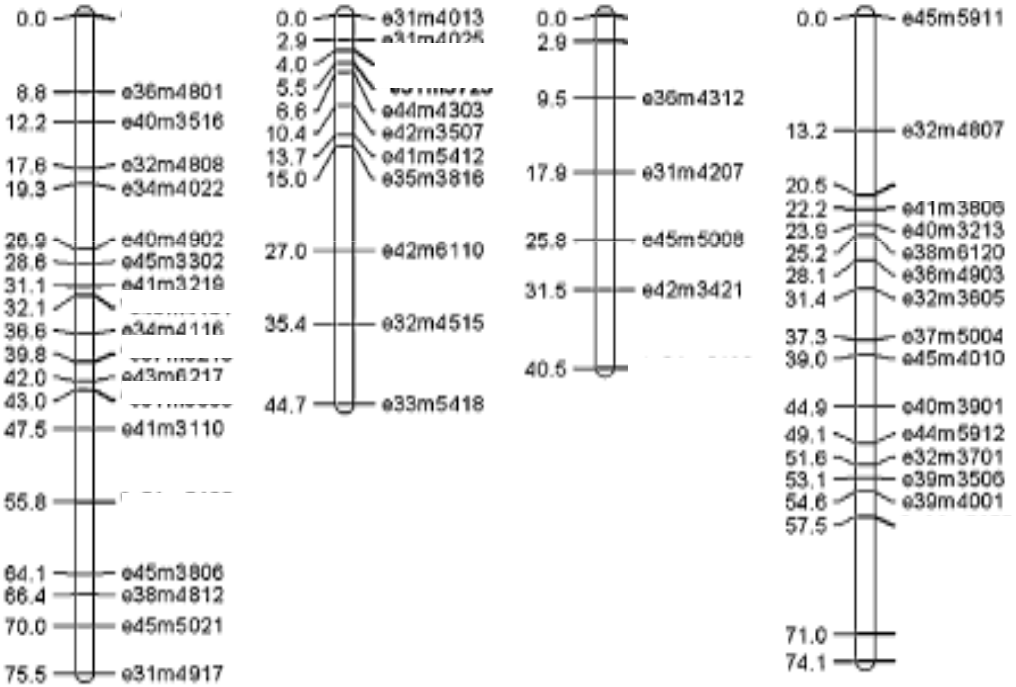
Table 2 Marker pair configurations and expected phenotypic frequencies used in this study

Hexasomic inheritance marker-pair configuration[a]	Phenotype	probabilities
Simplex/simplex coupling	AB	$1/2(1 - r)$
AB/00/00/00/00/00 × 00/00/00/00/00/00	A	$1/2r$
	B	$1/2r$
	0	$1/2(1 - r)$
Simplex/duplex coupling	AB	$1/2 - 1/5r$
AB/0B/00/00/00/00 × 00/00/00/00/00/00	A	$1/5r$
	B	$3/10 + 1/5r$
	0	$1/5 - 1/5r$
Duplex/duplex coupling	AB	$4/5 - 2/5r + 1/5r^2$
AB/AB/00/00/00/00 × 00/00/00/00/00/00	A	$2/5r - 1/5r^2$
	B	$2/5r - 1/5r^2$
	0	$1/5 - 2/5r + 1/5r^2$
Simplex/triplex coupling	AB	$1/2 - 1/20r$
AB/0B/0B/00/00/00 × 00/00/00/00/00/00	A	$1/20r$
	B	$9/20 + 1/20r$
	0	$1/20 - 1/20r$
Duplex/triplex coupling	AB	$4/5 - 1/4r + 1/20r^2$
AB/AB/0B/00/00/00 × 00/00/00/00/00/00	A	$1/10r - 1/20r^2$
	B	$3/20 - 1/10r - 1/20r^2$
	0	$1/20 - 1/10r + 1/20r^2$

[a] Distribution of alleles of two loci in a base chromosome group (chromosomes are separated by "/"). "A": presence of band at locus A, "B": presence of band at locus B, "0": absence of band. (*Source*: Kriegner et al. 2003; Ripol et al. 1999)

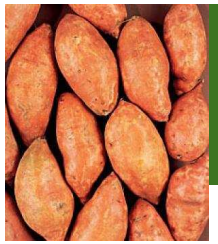


Mapping in autopolyploids



One linkage group per chromosome (90) of each parent => in total ~180 linkage groups

How to know which linkage groups are homologous?



Mapping in autopolyploids

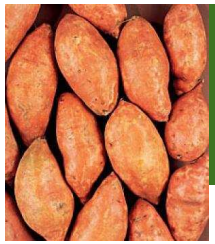
ESTIMATION OF RECOMBINATION FREQUENCIES

Multiplex (duplex and triplex) markers allow to establish homology relationships among the 6 parental chromosomes

Table 2 Marker pair configurations and expected phenotypic frequencies used in this study

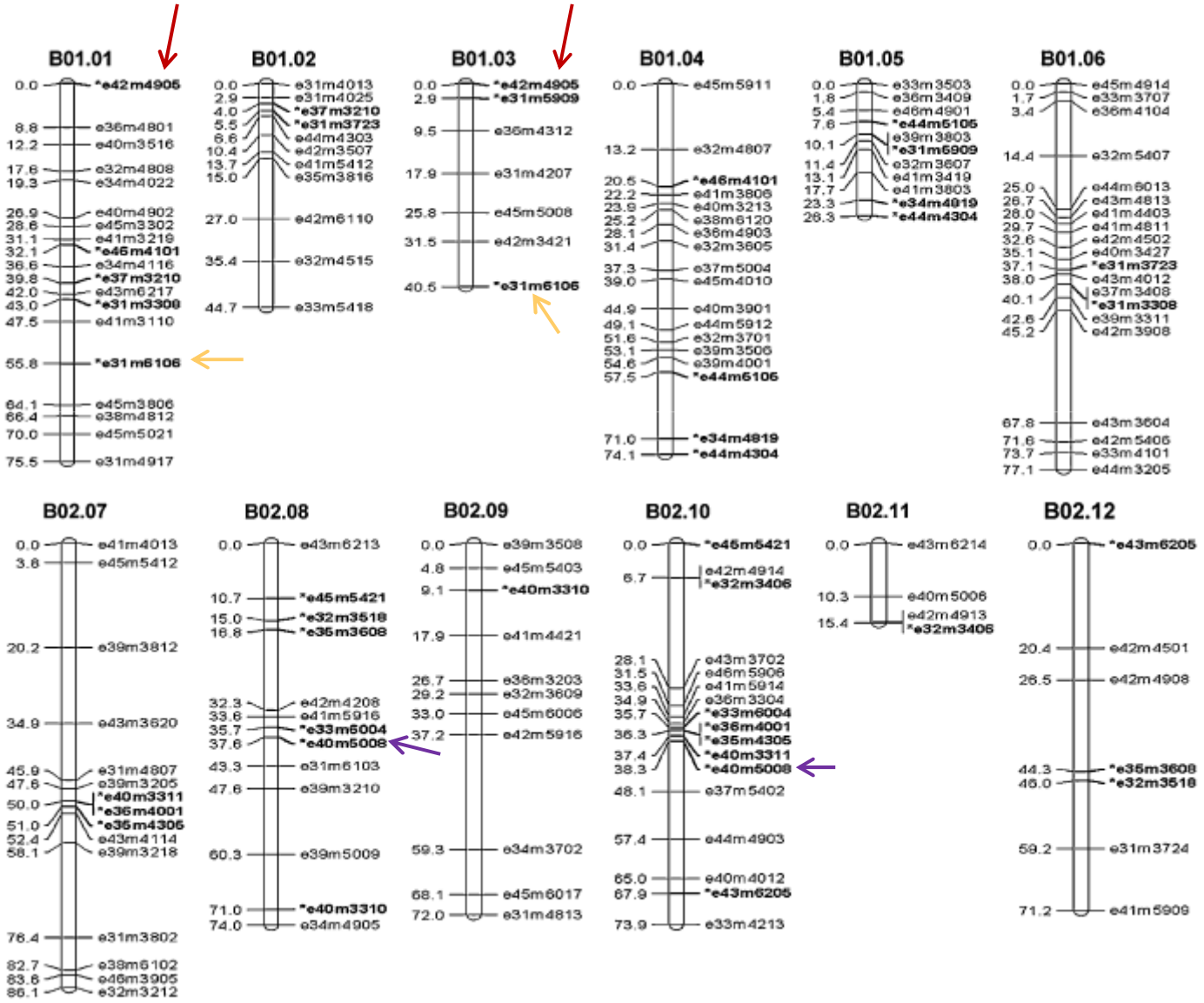
Hexasomic inheritance marker-pair configuration[a]	Phenotype probabilities	
Simplex/simplex coupling	AB	$1/2(1 - r)$
AB/00/00/00/00/00 × 00/00/00/00/00/00	A	$1/2r$
	B	$1/2r$
	0	$1/2(1 - r)$
Simplex/duplex coupling	AB	$1/2 - 1/5r$
AB/0B/00/00/00/00 × 00/00/00/00/00/00	A	$1/5r$
	B	$3/10 + 1/5r$
	0	$1/5 - 1/5r$
Duplex/duplex coupling	AB	$4/5 - 2/5r + 1/5r^2$
AB/AB/00/00/00/00 × 00/00/00/00/00/00	A	$2/5r - 1/5r^2$
	B	$2/5r - 1/5r^2$
	0	$1/5 - 2/5r + 1/5r^2$
Simplex/triplex coupling	AB	$1/2 - 1/20r$
AB/0B/0B/00/00/00 × 00/00/00/00/00/00	A	$1/20r$
	B	$9/20 + 1/20r$
	0	$1/20 - 1/20r$
Duplex/triplex coupling	AB	$4/5 - 1/4r + 1/20r^2$
AB/AB/0B/00/00/00 × 00/00/00/00/00/00	A	$1/10r - 1/20r^2$
	B	$3/20 - 1/10r - 1/20r^2$
	0	$1/20 - 1/10r + 1/20r^2$

[a] Distribution of alleles of two loci in a base chromosome group (chromosomes are separated by "/"). "A": presence of band at locus A, "B": presence of band at locus B, "0": absence of band. (*Source*: Kriegner et al. 2003; Ripol et al. 1999)



Mapping in autopolyploids

Keep order of simplex markers as ‘fixed’, and try to ‘place’ multiplex markers int he map



Software available

Diploids and allopolyploids:

- JoinMap (http://www.kyazma.nl)
- MapMaker
- Carte Blanche (for extremely large datasets)….

Autotetraploids:

- Tetraploidmap

Autohexaploids:

- Software for diploids, but using strategy described above

Use of linkage maps

- Locate genes affecting trait of interest (statistical associations between DNA-marker variants and trait)
  - Genetically simple traits (monogenic); disease resistances in plants are frequently controlled by one or a few genes
  - Genetically complex traits, involving many genes (quantitative trait loci, QTL), and affected by the environment; economically important traits such as yield or stress resistances
- Fine-mapping of genes
- Compare the genomes of different species
- Select DNA-markers evenly spread over the genome to perform a diversity study

BSA-analysis

Institute for Agricultural and Fisheries Research
Plant Sciences Unit
www.ilvo.vlaanderen.be
Agriculture and Fisheries Policy Area

BSA

DEFINITION

Bulked: it makes use of bulked samples

=> saves time and money

Segregant: it makes use of a segregating population

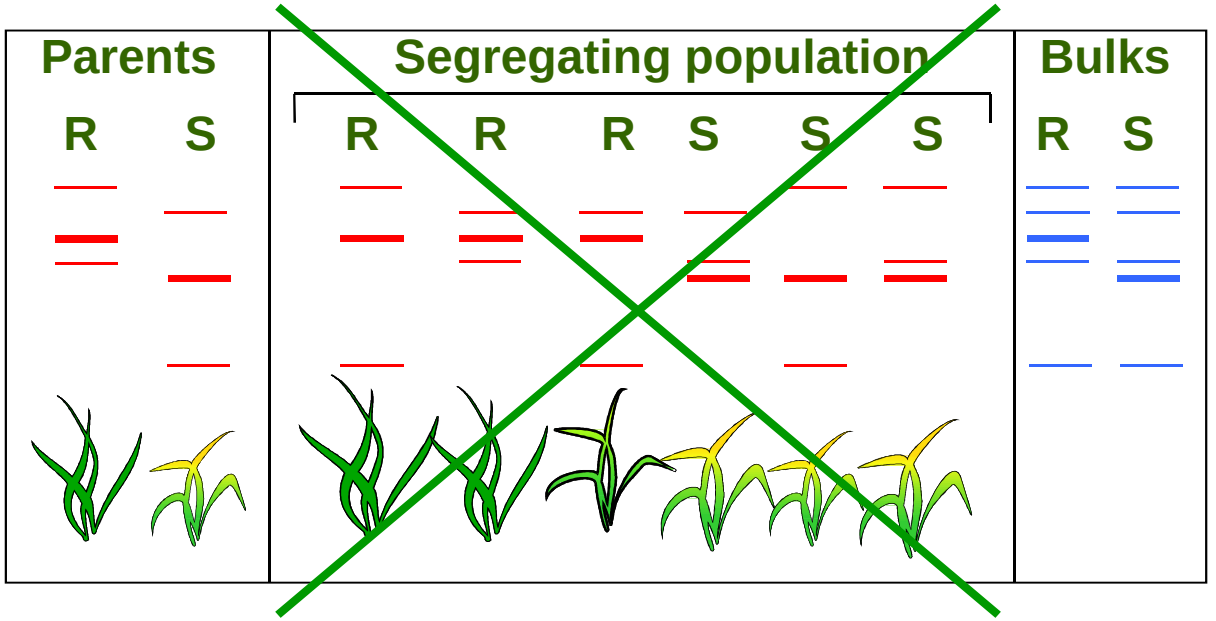
=> but it does not require map information!!!

Analysis: it screens the whole genome

BSA

DEFINITION

- Involves comparing two pooled DNA samples of individuals from a segregating population originating from a single cross
- Within each bulk the individuals are identical for the loci of interest but are arbitrary for all other loci
- Markers that are polymorphic between the pools are markers putatively linked to the loci involved in the trait of interest



BSA

DEFINITION

- BSA assumes that markers adjacent to (=linked to) the target locus will be in linkage disequilibrium (LD*) among themselves and with respect to the trait
  - Recombination will not have randomised these markers with respect to the targeted locus
  - As genetic distance increases, more recombinants will be present in each bulk, culminating in 50% recombinants (= no linkage disequilibrium) and therefore resulting in no differences between the bulks

*** non-random association of alleles at different genomic loci**

*** non-random association of a marker-allele at a given locus with a phenotype**

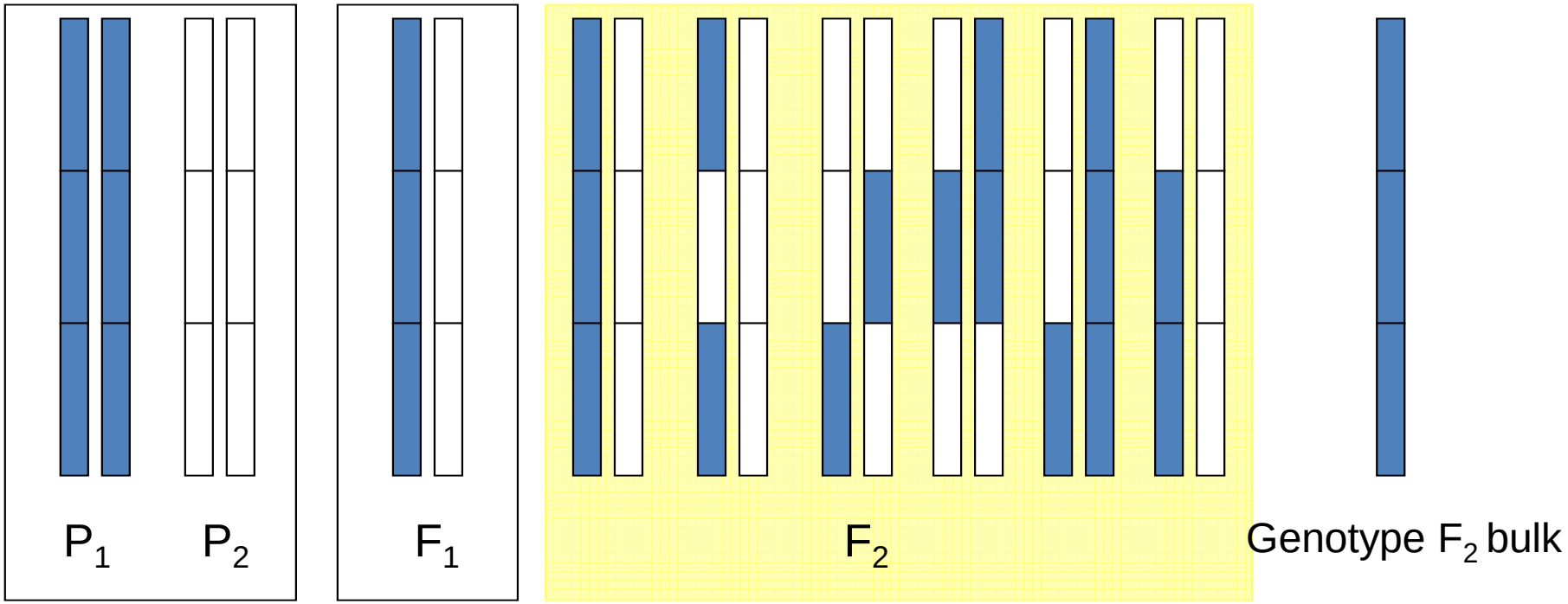
*** related to Mendel's second Law**

ILVO

BSA

DEFINITION

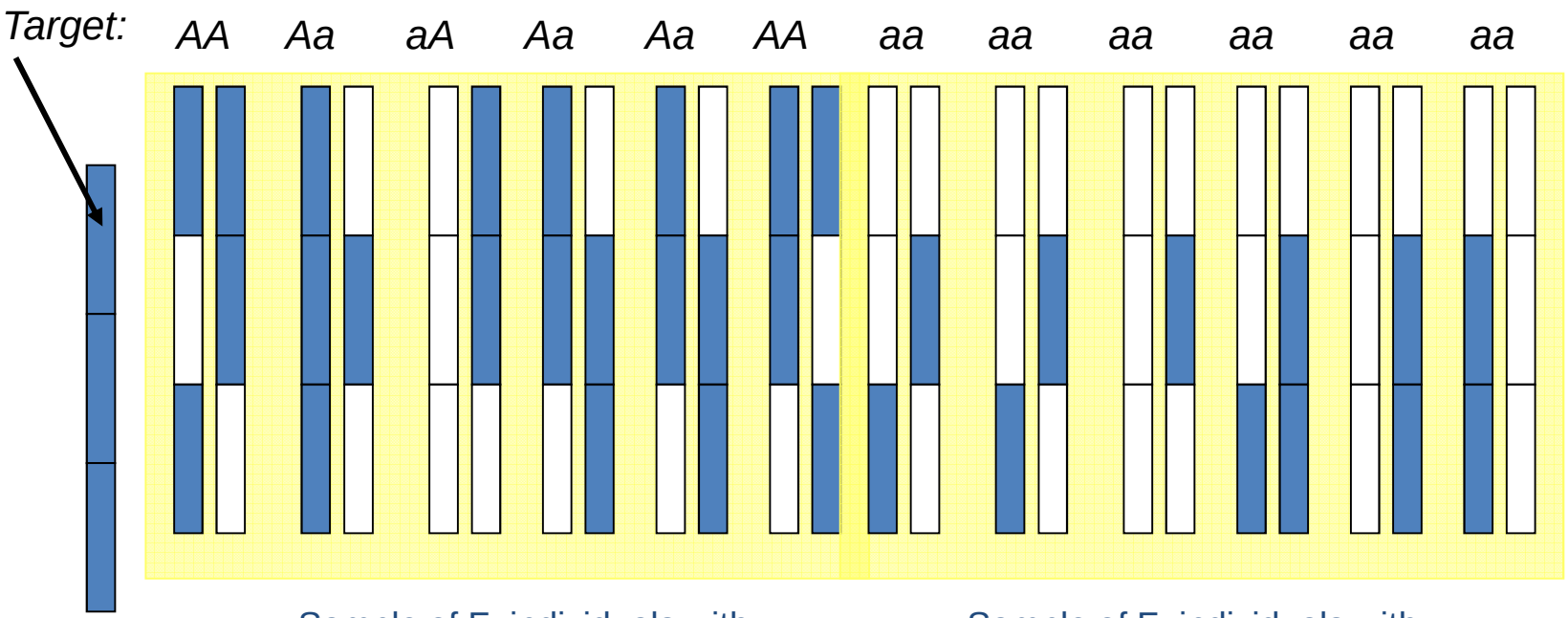
At a locus not linked to the trait, using dominant markers



BSA

DEFINITION

At a locus linked to the trait, using dominant markers



Sample of F_2 individuals with dominant phenotype harboring the selected locus:
this bulk is made using only phenotypic information and homozygote dominants cannot be distinguished from heterozygotes

Genotype bulk

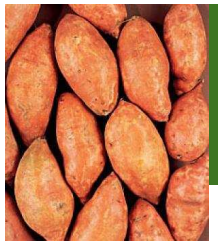
Sample of F_2 individuals with recessive phenotype harboring the selected locus:
only the homozygote recessive will be included in this bulk, made using only phenotypic information

Genotype bulk

BSA

STEPS

1) Create a segregating population from a single cross (for example, F2 or BC progeny)
2) Phenotype the progeny and identify individuals with extreme trait-phenotypes
3) Construct DNA bulks of the individuals displaying the most extreme trait-phenotypes
4) Genotype the parents and the bulks using hundreds to thousands of DNA-markers
5) Identify those markers which distinguish the bulks and the parents



BSA

EXAMPLE SWEETPOTATO

Ukoskit et al. (1997) Identifying a randomly amplified polymorphic DNA (RAPD) marker linked to a gene for root-know nematode resistance in sweetpotato

Fenotype screening

- F1 progeny of a single cross => log total nematode number => qualitative trait

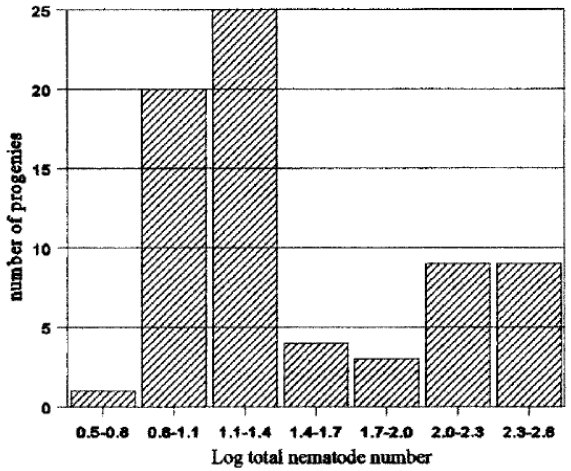
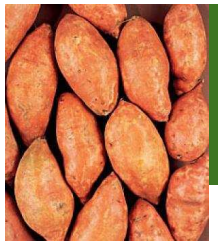


Fig 1. Frequency distribution of F_1 single-cross progenies for log total nematode number.

- Segregation in progeny fits 4:1 ratio => RRrrrr x rrrrrr



BSA

EXAMPLE SWEETPOTATO

Ukoskit et al. (1997) Identifying a randomly amplified polymorphic DNA (RAPD) marker linked to a gene for root-know nematode resistance in sweetpotato

Genotype screening

- R bulk, S bulk, R parent, S parent

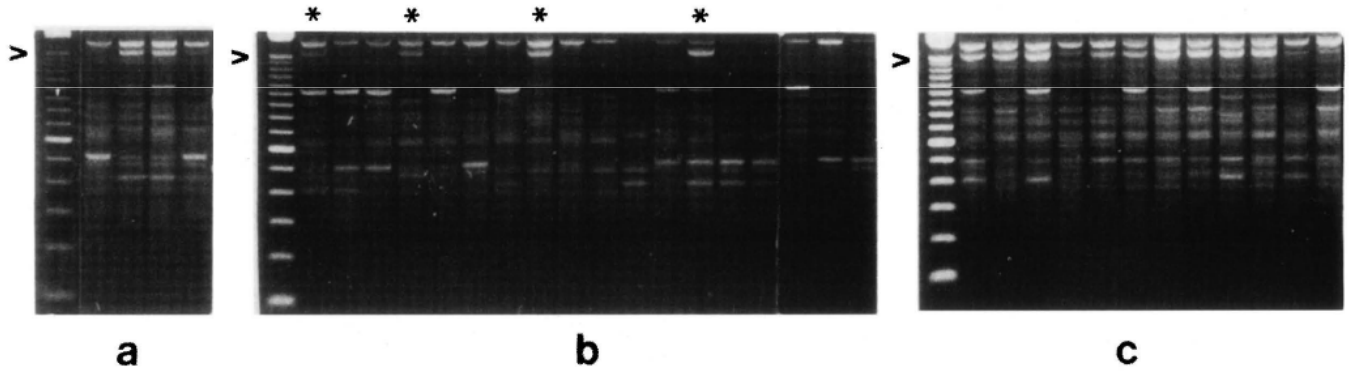


Fig 2. Bulk segregant analysis of $OPI5_{1500}$. (**a**) Lane 2 = susceptible parent, lane 3 = resistant parent, lane 4 = resistant bulk, lane 5 = susceptible bulk. (**b**) All susceptible progenies (* = recombinant progenies, susceptible with marker). (**c**) Example of resistant progenies. Arrows indicate the polymorphic band at 1500 bp. The first lane of each picture shows 100-bp molecular mass marker.

- 728 polymorphic bands screened; one linked with resistance; linkage is not complete (recombinants present); recombination fraction = 0.2421

QTL-analysis

Institute for Agricultural and Fisheries Research
Plant Sciences Unit
www.ilvo.vlaanderen.be
Agriculture and Fisheries Policy Area

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

- A **QTL** is the location of a gene (or set of genes) that affects a trait that is measured on a quantitative scale. Examples of quantitative traits are plant height, fruit size, grain yield…
- These traits are typically affected by more than one gene, and also by the environment
- Mapping QTL is not as simple as mapping a single gene that affects a qualitative trait
- Tools required:
  - Many polymorphic molecular markers ordered in a linkage map
  - Variation for the trait in the segregating population
  - Detailed and accurate phenotypic data of the segregating population
  - Appropriate statistical tools

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

- Also based on the assumption that markers adjacent to (=linked to) the targeted loci will be in linkage disequilibrium with respect to the trait

but

- QTL-analysis tries to identify simultaneously the chromosomal location of all the genetic factors affecting the trait
- In a QTL analysis we infer the QTL genotypes in order to estimate the QTL effects and locations from associations with known markers
- A QTL is described by
  1. Its chromosomal location
  2. The magnitude of its phenotypic effect
  3. The effect of gene dosage at the locus
  4. Its interactions with other QTLs

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

Chromosomal location:

Derived, for example, from associations between marker genotypes and trait phenotypes

Phenotypic effects:

In most studies it is found that a few genes explain large proportions of the phenotypic variance, with increasing numbers of genes explaining progressively smaller fractions of phenotypic variance

Effects of gene dosage:

The “gene action” at QTLs is determined by the same principles (additivity, dominance, recessivity…) as are employed for monogenetic traits

Epistasis (= interactions between QTLs)

The collective activities of mapped QTLs explain only a portion of the phenotypic difference between parents. This is usually explained by the importance of epistasis, or nonlinear interactions between unlinked loci

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

- χ^2 test of association or simple linear regression (no linkage map required)
  - **Pro:**
    - robust to violations of normality in phenotypic data
  - **Con:**
    - Cannot extract all the information from the data
    - Large progenies required
    - Bad resolution of QTL position obtained
    - Multiple testing problem

    Currently only used for initial data exploration

- Single marker analysis
  - Tests for differences in the means of the genetic marker classes
  - Rough estimation of the location of a QTL

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

P1 : MQ/MQ x P2: mq/mq

⇩

F1 : MQ/mq x P1

⇩

Genotype	Recombinant	Exp. freq
MQ/MQ	NO	$(1-r_{MQ})/2$
MQ/Mq	YES	$r_{MQ}/2$
MQ/mQ	YES	$r_{MQ}/2$
MQ/mq	NO	$(1-r_{MQ})/2$

Possible marker genotypes and their phenotypic means:

MQ/MQ and MQ/Mq: $\mu_{MM} = (1-r_{MQ})\ \mu_1 + r_{MQ}\ \mu_{BC1}$

MQ/mQ and MQ/mq: $\mu_{Mm} = r_{MQ}\ \mu_1 + (1-r_{MQ})\ \mu_{BC1}$

r_{MQ} = recombination freq between marker and QTL

μ_1 = average trait value of the plants with marker genotype MM

μ_{BC1} = average trait value of the BC population

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

- Expected difference in average trait values between the two marker classes (MM and Mm) is:

  $\mu_{MM} - \mu_{Mm} = (1-2r_{MQ})\delta$

  δ = difference in trait values between P1 and P2

- If the trait and the marker are unlinked, $r_{MQ} = 0.5$:

  $\mu_{MM} - \mu_{Mm} = 0$

- What we have to do then is test the hypotheses:

  $H_0 : \mu_{MM} - \mu_{Mm} = 0$

  $H_1 : \mu_{MM} - \mu_{Mm} \neq 0$

- Standard t-test to compare the means will determine whether the marker genotype has any effect on the trait.
- If we conclude that H_1 is true => the marker has an effect on the trait, and we can conclude that the marker *M* is located in the neigbourhood of QTL *Q*.

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

Single marker analysis

Pro:

- Robust to violations of normality in phenotypic data

Con:

- The order of the M and the QTL on the genetic map remains unknown (MQ or QM?)
- Low power if the markers are far apart
- Large progenies required
- It is not possible to distinguish between size of a QTL effect and its position (relative to the marker): a marker close to a QTL of small effect will give the same 'signal' as a marker some distance from a QTL of large effect
- t-testing is not appropriate due to deviations from normality. Non-parametric tests such as Kruskal-Wallis are more appropriated
- Multiple testing problem

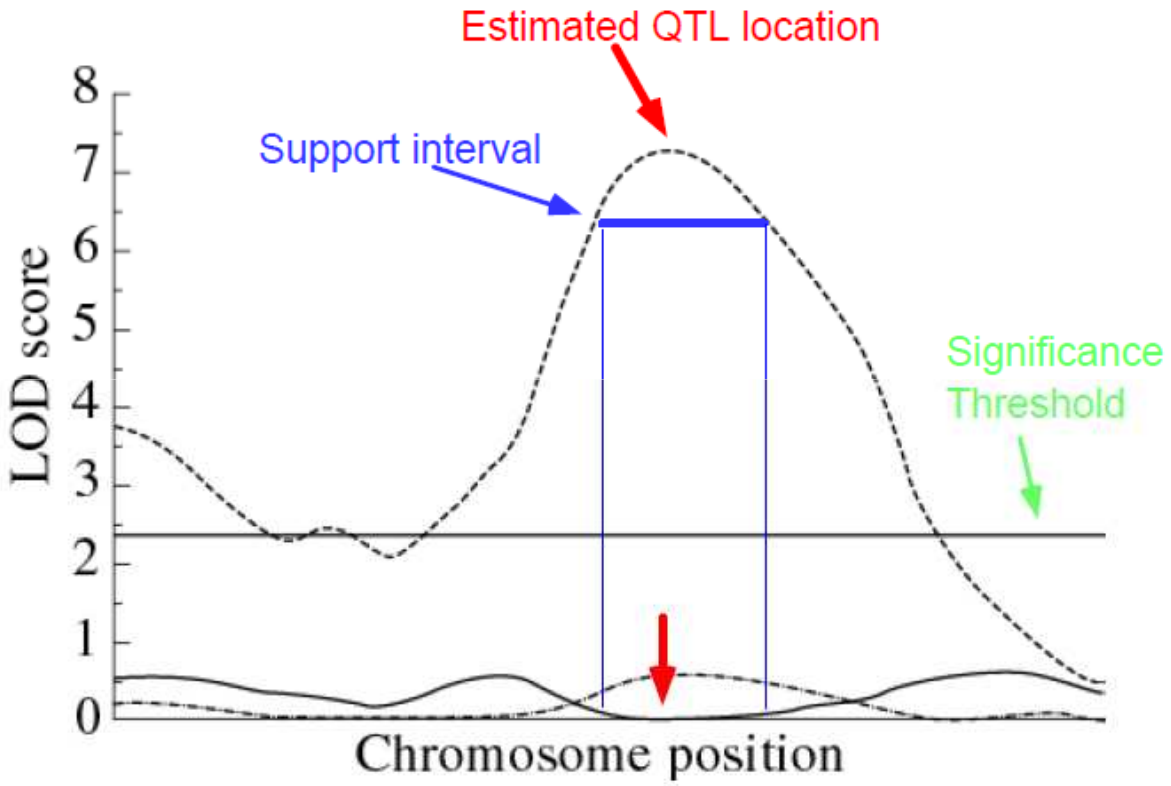
Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

- Interval mapping (requires linkage map)
  - Estimates the position of a QTL between two markers
  - Systematically searches the genome by calculating a test statistic at each position of the genome
  - Originally based on the maximum likelihood estimates
  - Intervals between adjacent markers along a chromosome are scanned and the LOD of there being one versus no QTL at a particular point is estimated
  - The higher the value of the LOD score, the more likely the data are if there was a QTL present compared to the situation when there is no QTL
  - A LOD-profile is constructed along the chromosome, and the maxima in this profile which exceed a specified significance level, indicate likely sites of a QTL
  - **Pro:**
    - By taking into account several markers simultaneously, it allows accurate estimation of QTL-positions
    - It is possible to separate the recombination and the size of the gene effect
    - Precision and power are increased by the use of extra information from a second marker
  - **Con:**
    - The effect of other QTLs present in the genome is neglected, and only the QTLs with the biggest phenotypic QTL on a chromosome, estimated positions and effect may be biased
    - Multiple testing problem

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING



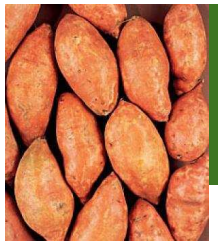
LOD = Log of the odds
$Z = \log_{10} (L1/L0)$
Interpretation= the alternative hypothesis (L1) is 10^Z times more likely than the null hypothesis (L0)

Quantitative traits

QUANTITATIVE TRAIT LOCUS (QTL) MAPPING

- Composite Interval mapping (CIM) or Multiple QTL mapping (MQM)
  - Markers located nearby putative QTLs identified by e.g. IM, are used as co-factors in an approximate multiple-QTL model. At each testing point the effect of one or more co-factors is included.
  - By entering QTLs identified by IM (with the biggest effects) as co-factors, the effects of these QTLs is absorbed, increasing the power to identify additional QTLs.

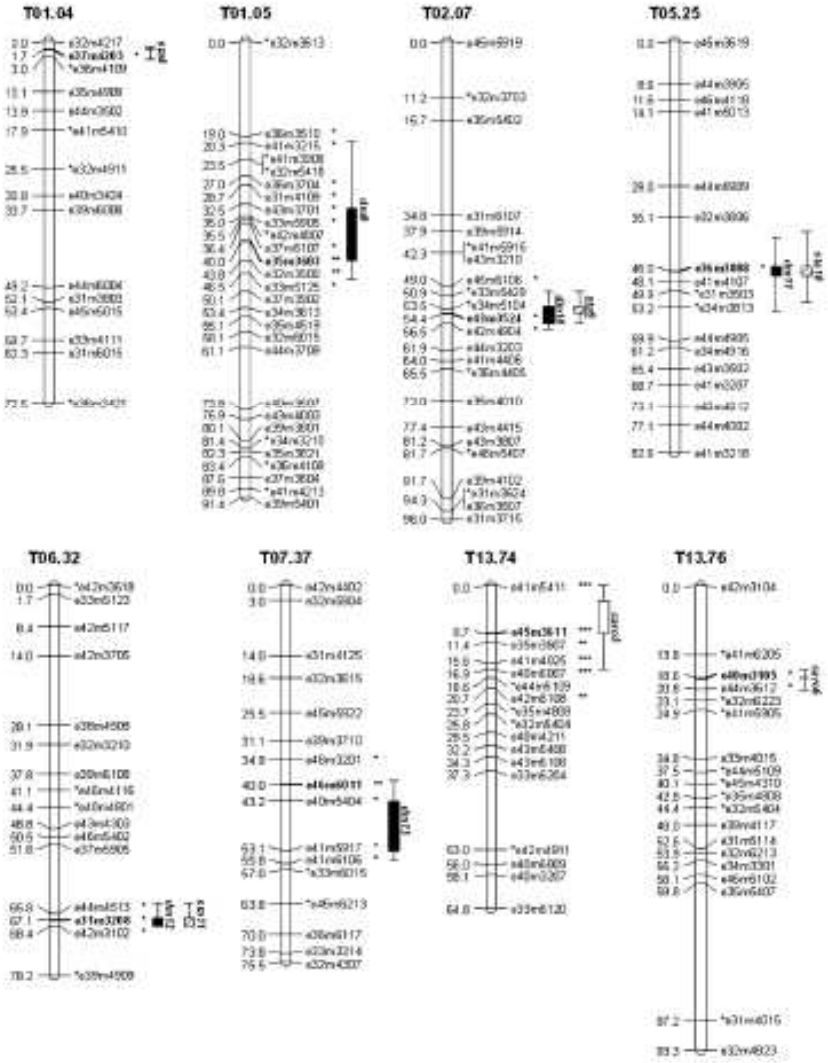
The most used method at present



Quantitative traits

QTL MAPPING (CIM)

Fig. 5 Tanzania linkage groups showing significant QTL for β-carotene content, starch content, and dry-matter content. Significance at 5, 1 and 0.1% is indicated by *, ** and ***, respectively. The *markers* most significantly associated with the trait, according to ANOVA analysis, are shown in *bold*. QTL are shown as *vertical bars* on the *right side* of the respective linkage group. *Shaded boxes* QTL are named and according to the trait they affect (dm = dry-matter content; sta = starch content; and caro = carotene content). QTL with positive effect are indicated by *normal text*, and those with negative effect are indicated by *underlined text*



Cervantes-Flores et al (2011)

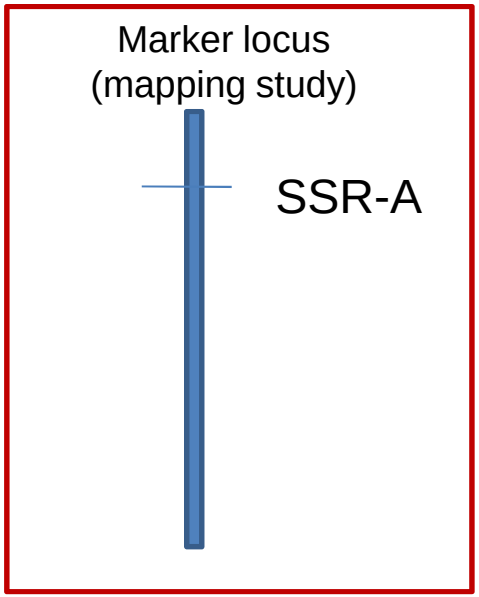
7. Developing markers for application

DNA-Marker Validation

- When we map one trait in one segregating population, we are analyzing the effects of the genetic factors carried by the parents of the mapping population, but we do not know anything about the situation in the ‘global population’ (outside the mapping population)
  - For example, in a mapping population we find that allele 1 at locus A (A1) is associated with disease resistance. We do not know yet whether in a different plant (not used for mapping) allele 1 at locus A will also be associated with disease resistance. In many cases it will be a different allele
  - A different genetic locus might be more relevant in other plants
- Test marker trait association in alternative populations => estimate reliability of marker in predicting phenotype
- Identify polymorphisms between lines used in breeding programs (perhaps different alleles present than in the mapping population)
- Develop a palette of suitable markers wih associated polymorphism data
  - E.g. 10 markers within 10 cM of trait for Marker Assisted Backcross selection (MAB)
  - Provide protocols and polymorphism data to breeder

7. Developing markers for application

DNA-Marker Validation



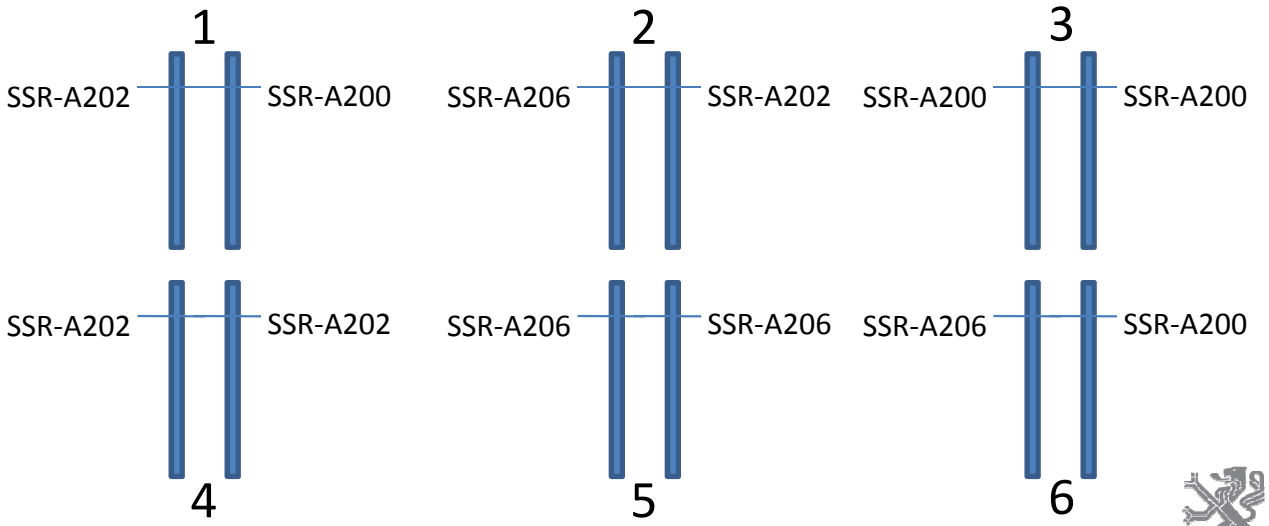
From mapping study:

- We know that marker SSR-A is linked to the gene of interest (e.g. disease resistance)
- We also know that allele SSR-A200, coming from parent P1 of the mapping population is associated with disease resistance

What can we say about plants 1-6 (not from mapping pop)?

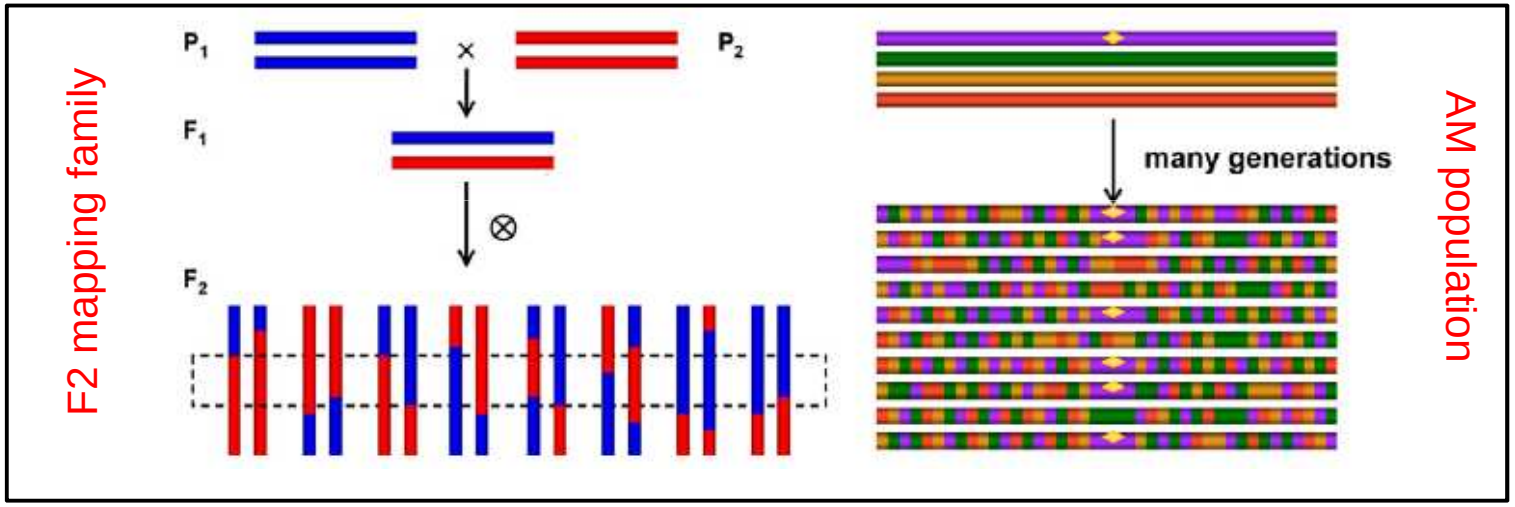
- They carry other alleles at locus SSR-A (e.g. SSR-A200, SSR-A202, SSR-A206, SSR-A210)
- Do we still find an association between allele SSR-A200 and resistance if plants 1, 3 and 6 are resistant and the rest susceptible?
- If plants 1, 2 and 4 are resistant and the rest susceptible?

Genetic information of 6 other plants for this locus



Association Mapping

Association analysis (= linkage disequilibrium mapping) or association mapping, is a population-based survey used to identify trait-marker relationships based on linkage disequilibrium (Flint-Garcia et al. 2003. Annu. Rev. Plant Biol. 54, 357-374)

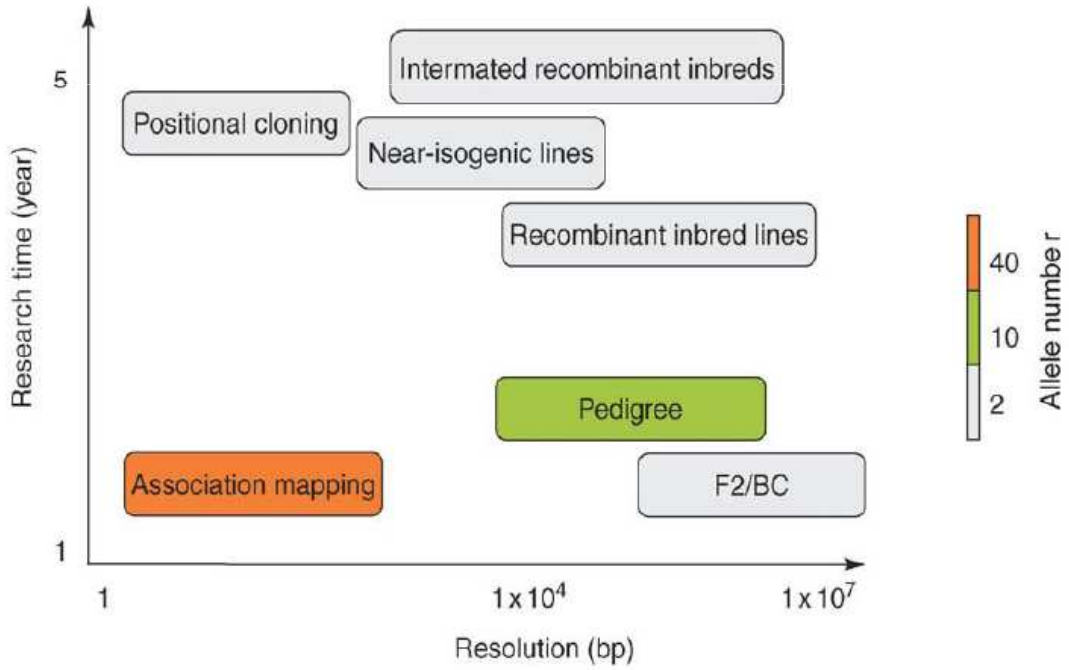


- F2 family and AM population: co-heritance (= co-segregation) of functional polymorphisms and neighboring DNA markers
- F2 family: only a few opportunities for recombination (only a few generations) => low mapping resolution
- AM population: historical recombination and natural genetic diversity exploited => high mapping resolution

Association Mapping

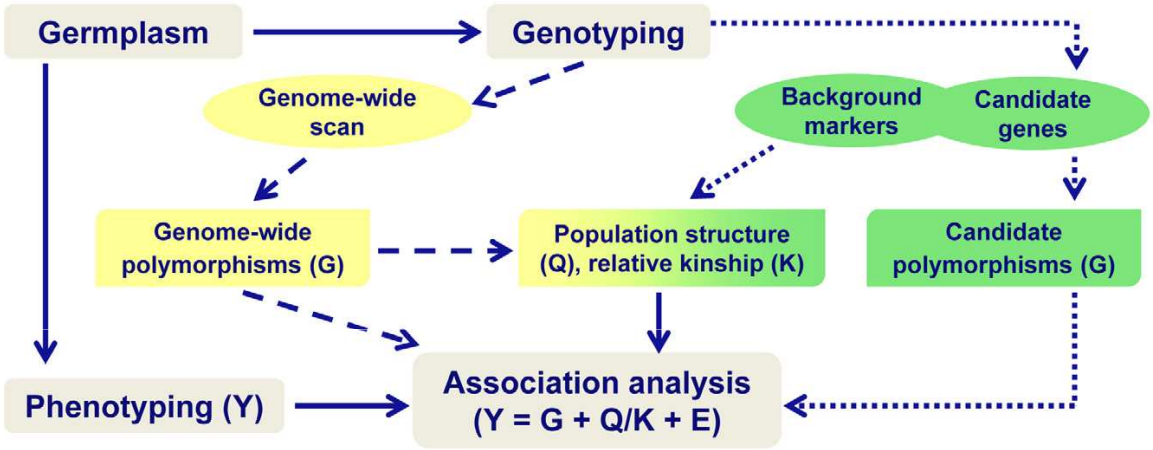
ADVANTAGES OF AM

- Increased mapping resolution
- Reduced research time
- Greater allele number
- Results readily applicable in breeding germplasm



Association Mapping

GENOME-WIDE / CANDIDATE-GENE



Genome-wide association mapping	Candidate-gene association mapping
It is a comprehensive approach to systematically search the genome for causal genetic variation. A large number of markers are tested for association with various complex traits, and prior information regarding candidate genes is not required. It works best for a research consortium with complementary expertise and adequate funding.	Candidate genes are selected based on prior knowledge from mutational analysis, biochemical pathway, or linkage analysis of the trait of interest. An independent set of random markers needs to be scored to infer genetic relationships. It is a low cost, hypothesis-driven, and trait-specific approach but will miss other unknown loci.

Q, K or both can be included in the analysis, depending on the genetic relationship of the association mapping population and the divergence for the trait examined

Q = population structure
K = relative kinship
E = residual variance

Association Mapping

REQUIREMENTS

- Natural diversity
  - AM exploits natural diversity and its results are applicable to a wide germplasm base
  - As QTLs are mapped in collections of breeding lines, landraces, or samples from natural populations, AM offers great potential for future trait improvement
- Genomic technology
  - Technology at low cost/data point (SNPs) – new sequencing technologies
  - Information on location and function of genes involved (for candidate-gene approach)
- Methodology for data analysis
  - To date, few efforts for AM tools specifically adapted to plants
  - Specificities of plants: geographical origins, local adaptation, and breeding history

ILVO