Molecular markers: an introduction

Isabel Roldán-Ruiz

August 2012

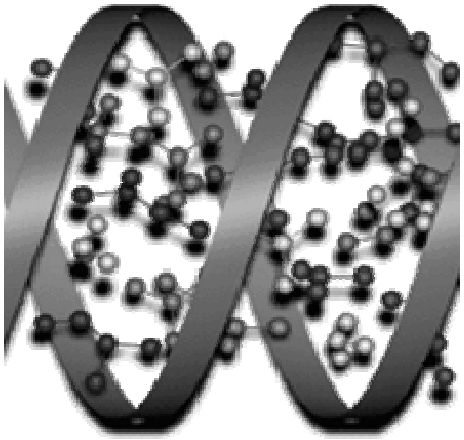
Institute for Agricultural and Fisheries Research
Plant Sciences Unit
www.ilvo.vlaanderen.be
Agriculture and Fisheries Policy Area

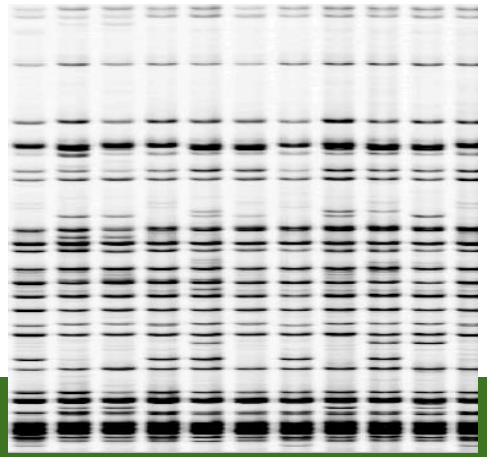
Contents

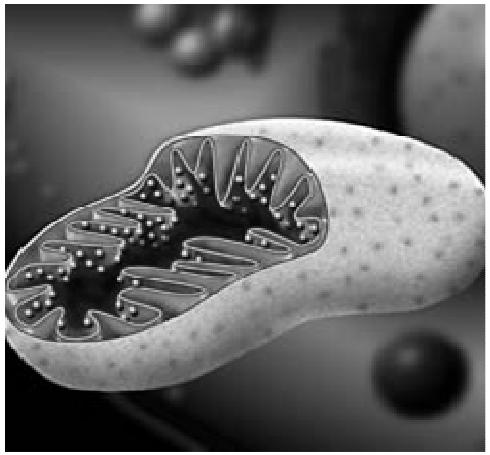
- Markers and breeding
- Some basic concepts
- Mutations, DNA-polymorphisms and DNA-markers
- DNA-marker techniques
- Some DNA-marker techniques used in the analysis of plant genomes: RAPD, SSR, CAPS, ISSR, AFLP, SNP

Markers and breeding

- Fields of application:
  - Markers linked to genes of interest are used to select in breeding populations
  - Marker-assisted introgression
  - Investigation of biological processes (mating system, pollen or seed dispersal)
  - Paternity analysis
  - Identification of specific genotypes (e.g. plant varieties)
  - Quantification of genetic diversity and relationships within and between (agricultural) species and populations



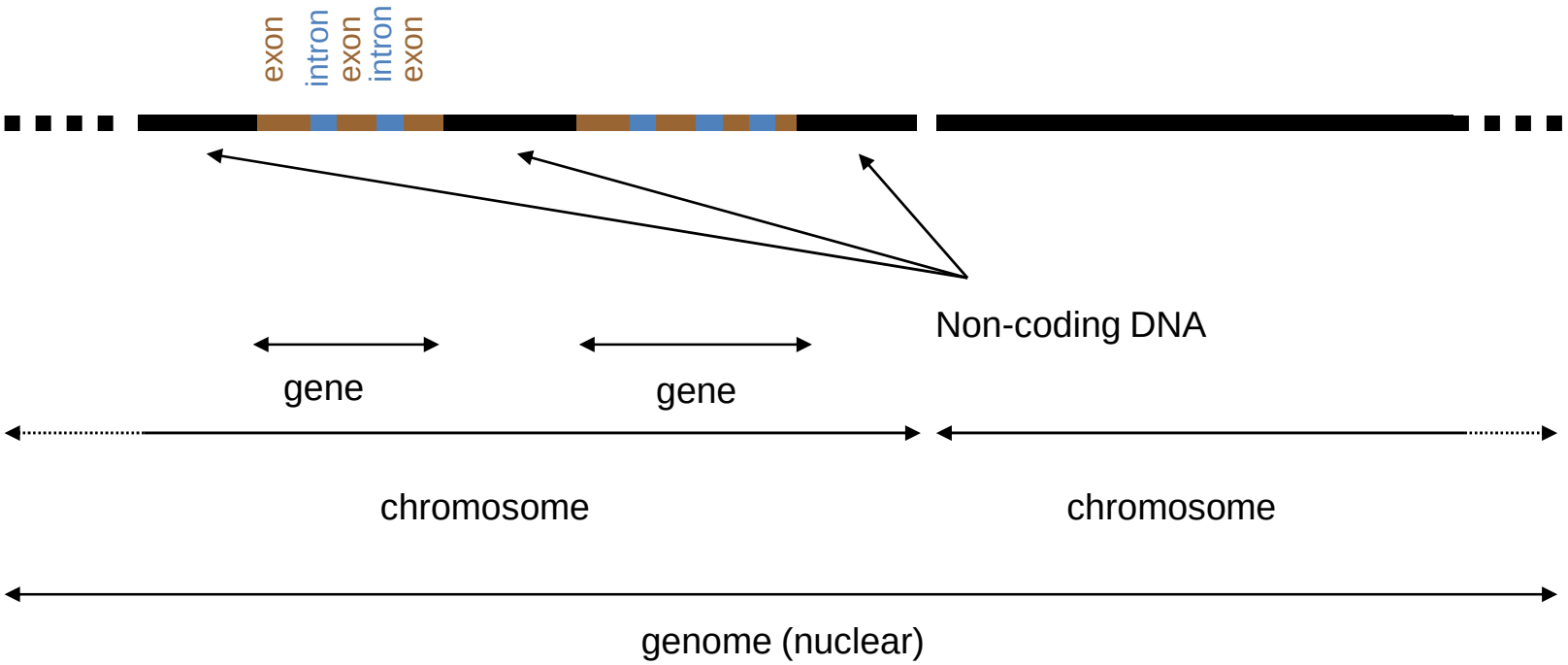




Some basic concepts

Some basic concepts

GENOME – CHROMOSOME - GENE



- Gene: certain region in the DNA molecule of the organism. A gene encodes in its sequence the potential to synthesize a particular protein or RNA molecule
- In eukaryotes, the DNA of the genome typically contains a large fraction that does not carry genes

Some basic concepts

MUTATION

- The primary requirement for natural or artificial selection is **genetic variability; MUTATION** generates genetic variability in the DNA strand, creating 'novel sequences'
- **Changes in the DNA-sequence and structure (= MUTATIONS) create variability**
- Mutations occur in coding regions and non-coding regions

1. **Point mutations:** due to errors in DNA-replication and DNA-repair
2. **Chromosomal sequence alterations**
    1. Translocations: nucleotide sequences are transferred from one chromosome to another
    2. Inversions: blocks of nucleotides rotate 180°
    3. Transpositions: nucleotide blocks move from place to place in the genome (transposons)
3. **Chromosomal additions and deletions**
    1. Deletions: chromosome blocks are lost within a chromosome
    2. Additions: nucleotide sequences are multiplied
4. **Chromosomal number changes**
    1. Aneuploidy: one or more chromosomes are absent or present in excess
    2. Haploidy: half of the normal chromosome set
    3. Polyploidy: more than two sets of homologous chromosomes

Some basic concepts

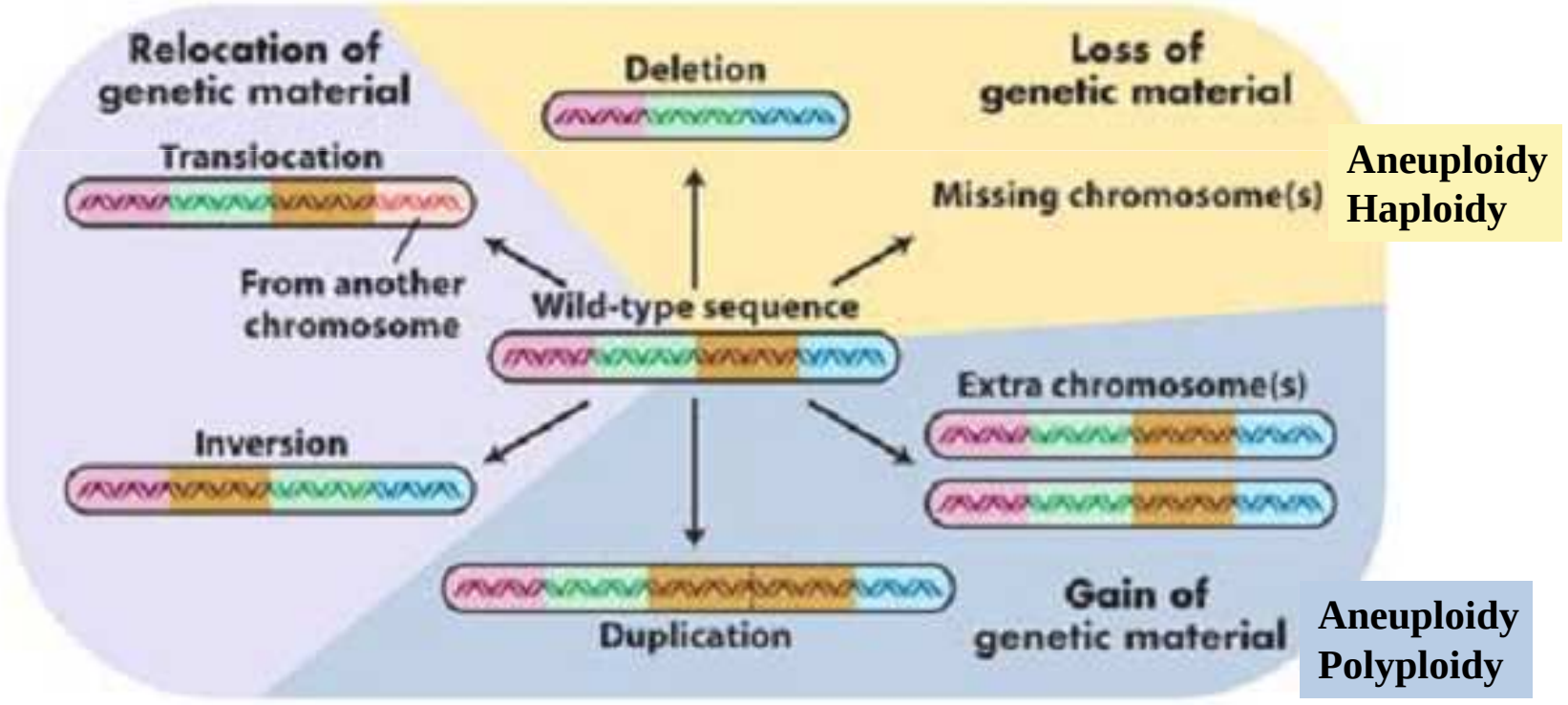
MUTATION TYPES

ACGTCAGTCACCAAA

↓

ACGTCAGGCACCAAA

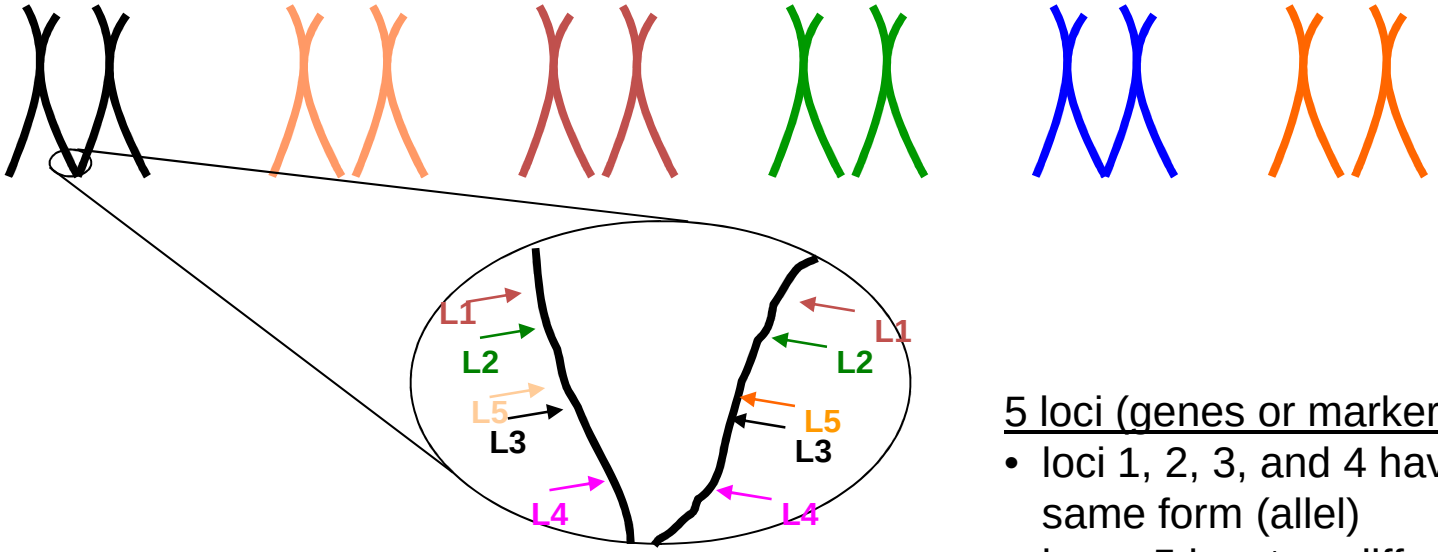
Point mutation



Some basic concepts

LOCUS – ALLELE – HOMO/HETEROZYGOUS

Diploid plant with 6 chromosomes (x= 6; 2n = 2x = 12)



5 loci (genes or markers):

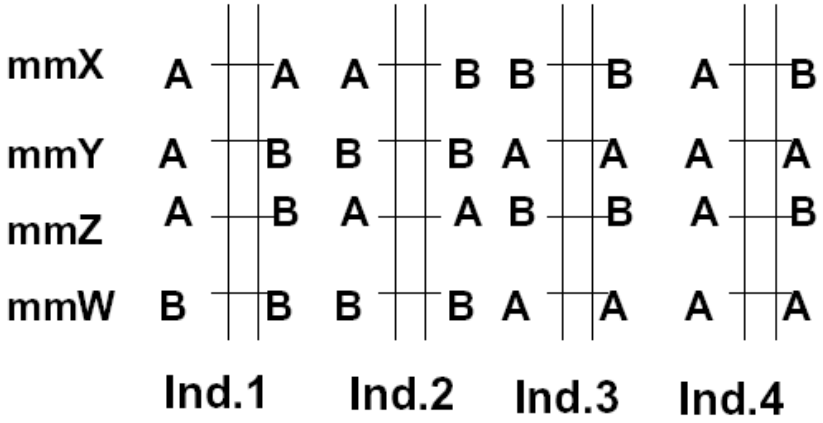
- loci 1, 2, 3, and 4 have the same form (allel)
- locus 5 has two different forms (alleles)

- In diploid plant cells, two copies of each chromosome are present. In plants polyploids are frequently found: each chromosome is present in more than two copies.
- In a diploid plant the two homologous chromosomes are very similar but not necessarily identical in DNA sequence.
- **Sweetpotato is hexaploid (2n = 6x = 90)**

Some basic concepts

DOMINANT – RECESSIVE - CODOMINANT

	mmX	mmY	mmZ	mmW
ind.1	AA	AB	AB	BB
ind.2	AB	BB	AA	BB
ind.3	BB	AA	BB	AA
ind.4	AB	AA	AB	AA



Genotypic frequencies:

	AA	AB	BB
mmX	0.25	0.5	0.25
mmY	0.5	0.25	0.25
mmZ	0.25	0.5	0.25
mmW	0.5	0	0.5

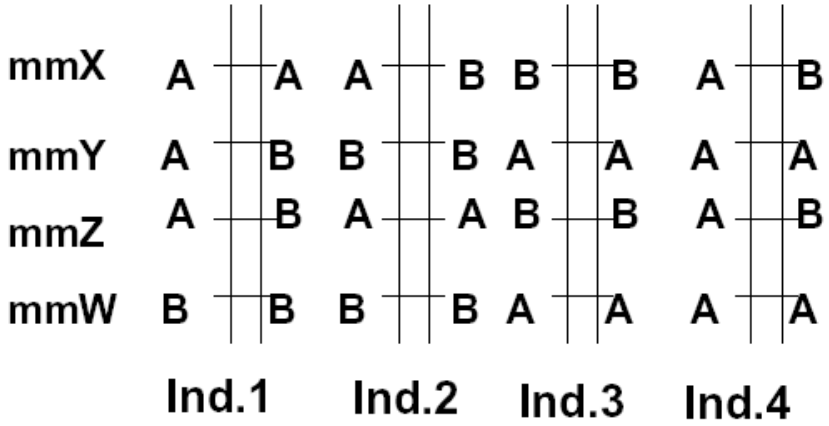
Allelic frequencies:

	A	B
mmX	0.5	0.5
mmY	0.375	0.625
mmZ	0.5	0.5
mmW	0.5	0.5

Some basic concepts

DOMINANT – RECESSIVE - CODOMINANT

	mmX	mmY	mmZ	mmW
ind.1	1	1	1	0
ind.2	1	0	1	0
ind.3	0	1	0	1
ind.4	1	1	1	1



Genotypic frequencies:

	AA	AB	BB
mmX	?	?	0.25
mmY	?	?	0.25
mmZ	?	?	0.25
mmW	?	?	0.5

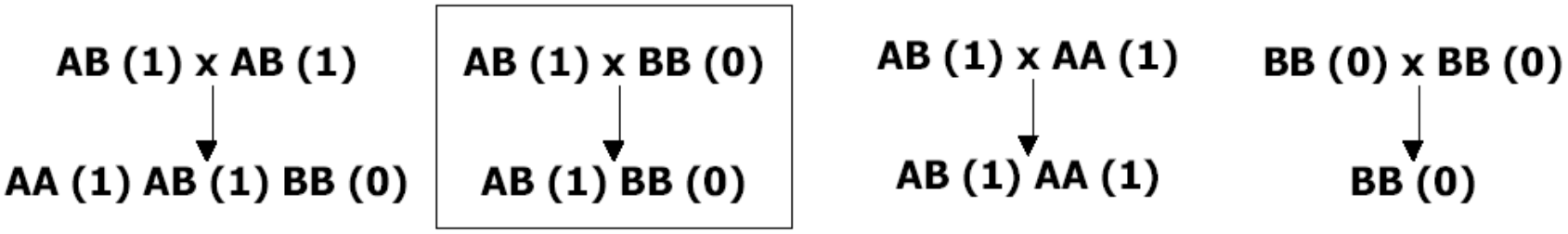
Allelic frequencies:

	A	B
mmX	?	?
mmY	?	?
mmZ	?	?
mmW	?	?

Some basic concepts

DOMINANT – RECESSIVE - CODOMINANT

impossible to determine whether a given plant is homozygous dominant (AA) or heterozygous (AB) for a given marker
BUT
if we know the state of the marker in the parents …



In backcross designs if the dominant allele is heterozygous it is possible to predict the genotype of the offspring

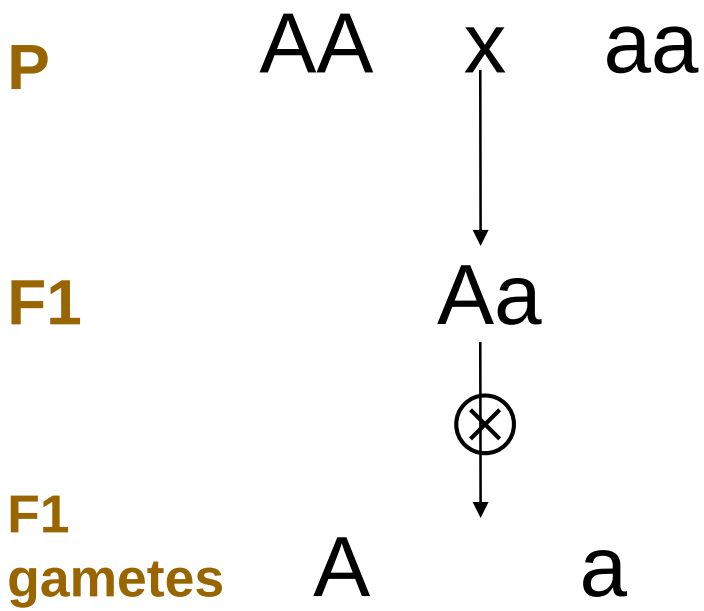
Some basic concepts

MENDEL'S FIRST LAW

During gamete formation each member of the allelic pair (= locus) separates from the other member to form the genetic constitution of the gamete

Some basic concepts

MENDEL'S FIRST LAW



Union of gametes at random:

	A	**a**
A	AA	Aa
a	Aa	aa

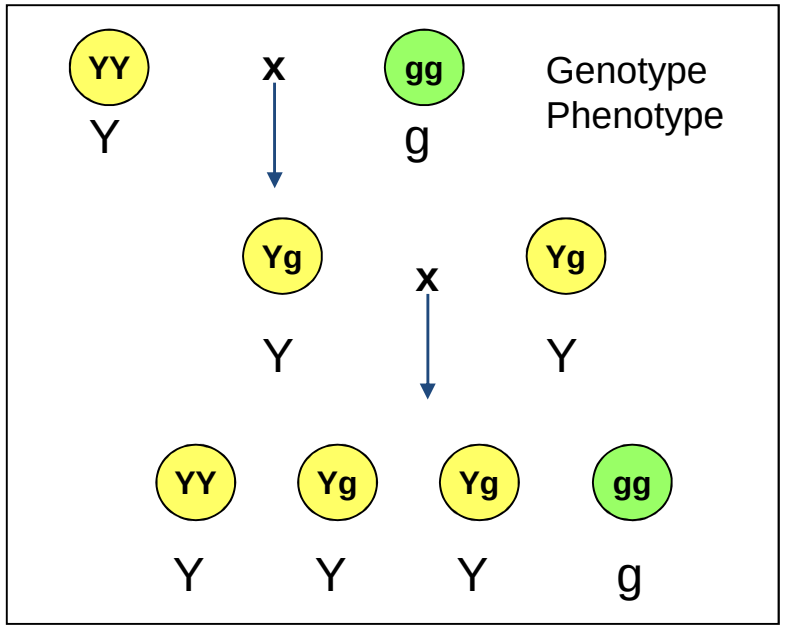
F2:

1/4AA : 1/2Aa : 1/4aa

Some basic concepts

MENDEL'S FIRST LAW

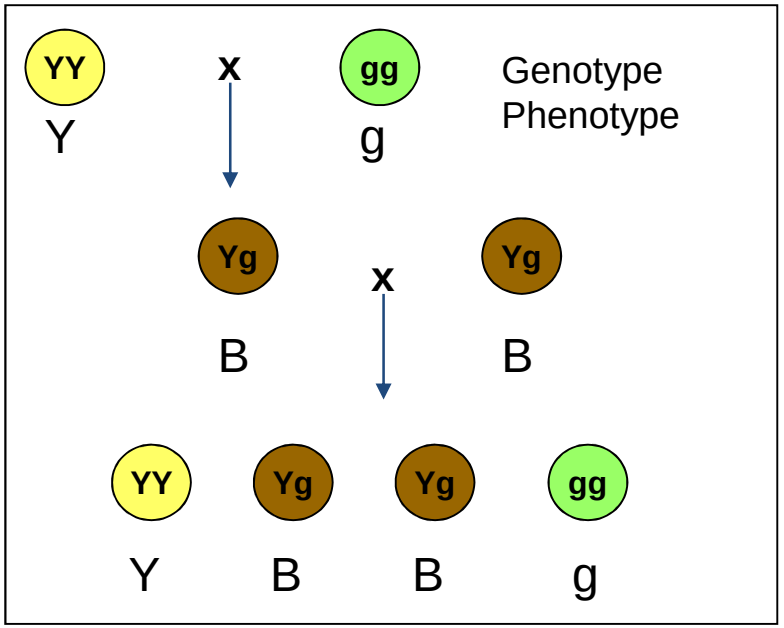
Mendel studied major gene differences in the garden pea: color of the seed (one gene with alleles Y and g, Y>g)



Dominant allele : Y

Recessive allele: g

Another possible outcome of this experiment (in other system, of course)

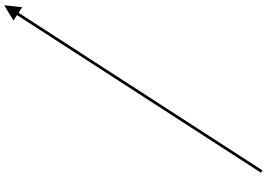


Y and g are co-dominant alleles

Some basic concepts

MENDEL'S SECOND LAW

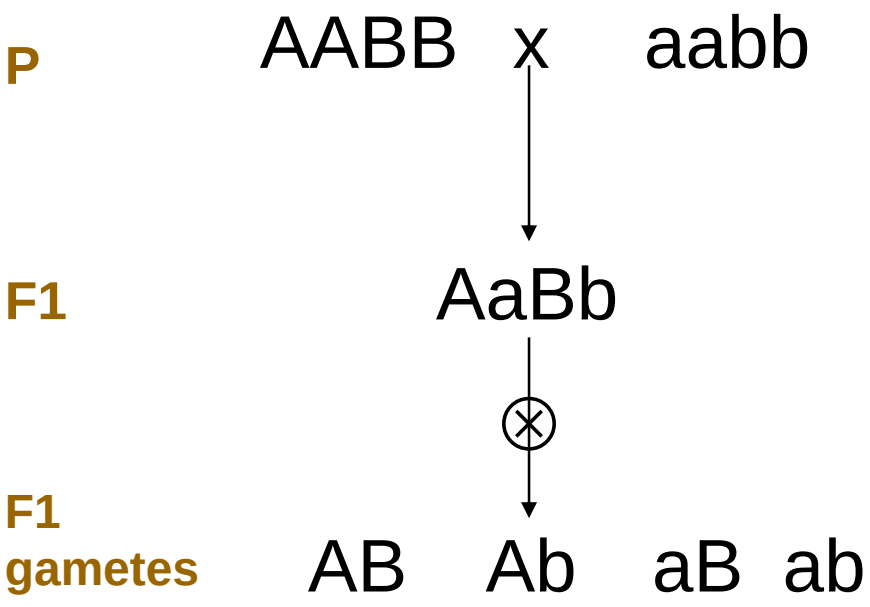
During gamete formation the segregation of the alleles of one allelic pair (= locus / gene) is independent of the segregation of the alleles of another allelic pair (= locus / gene)



This is the basic concept to understand linkage relationships among genes, among markers, and among genes and markers and to construct linkage maps

Some basic concepts

MENDEL'S SECOND LAW



Union of gametes at random:

	AB	**Ab**	**aB**	**ab**
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	Aabb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

F2:

9A_B_ : 3A_bb : 3aaB_ : 1aabb

Some basic concepts

MENDEL'S SECOND LAW

– During gamete formation the segregation of the alleles of one allelic pair (=locus / gene) is independent of the segregation of the alleles of another allelic pair (= locus / gene)

This is not always true: if the two loci are 'located close to each other' along the chromosome (= linked), they will not segregate independently

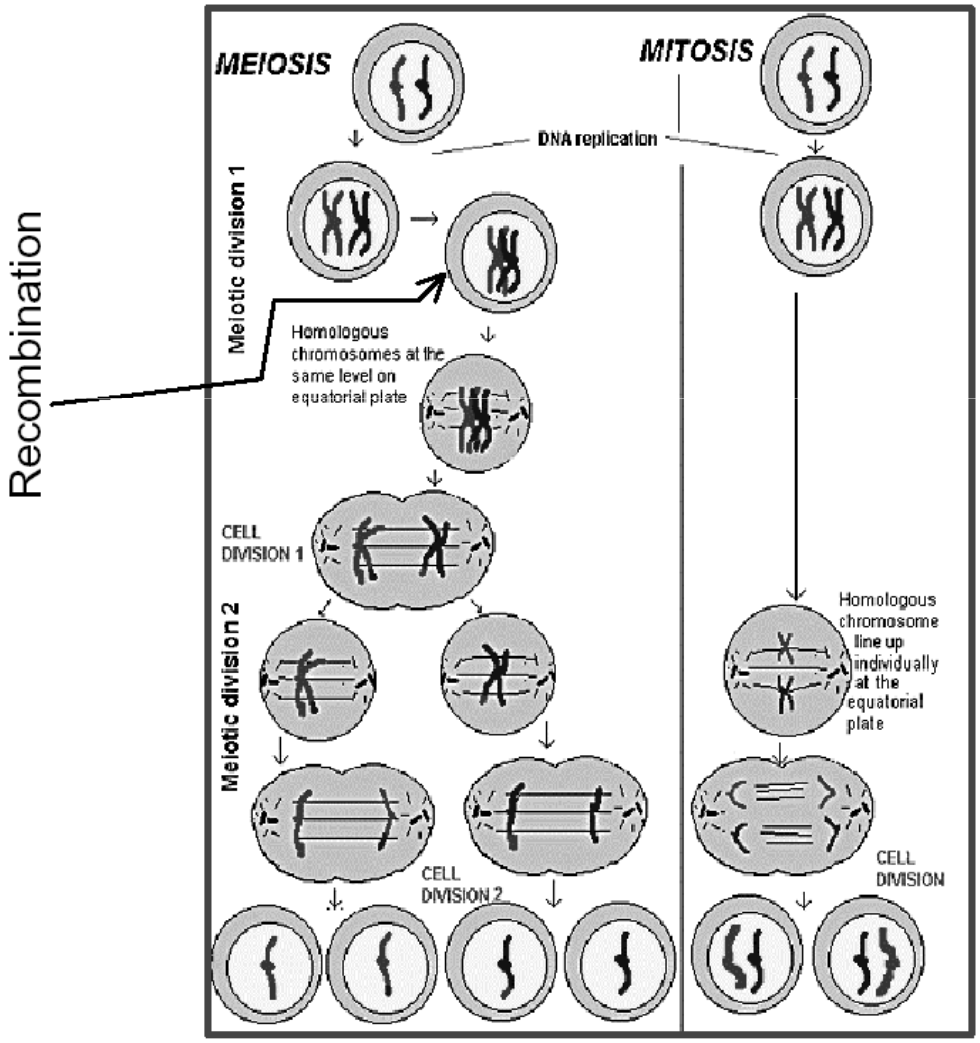
⇒ For linked loci, the observed segregation ratios deviate from the expectation
⇒ This information can be used to identify loci which are linked to each other and to determine the (recombination) distance among them (the closer the linkage the lower the recombination frequency and the larger the deviation from the expectation)

Some basic concepts

LINKAGE AND RECOMBINATION

Recombination

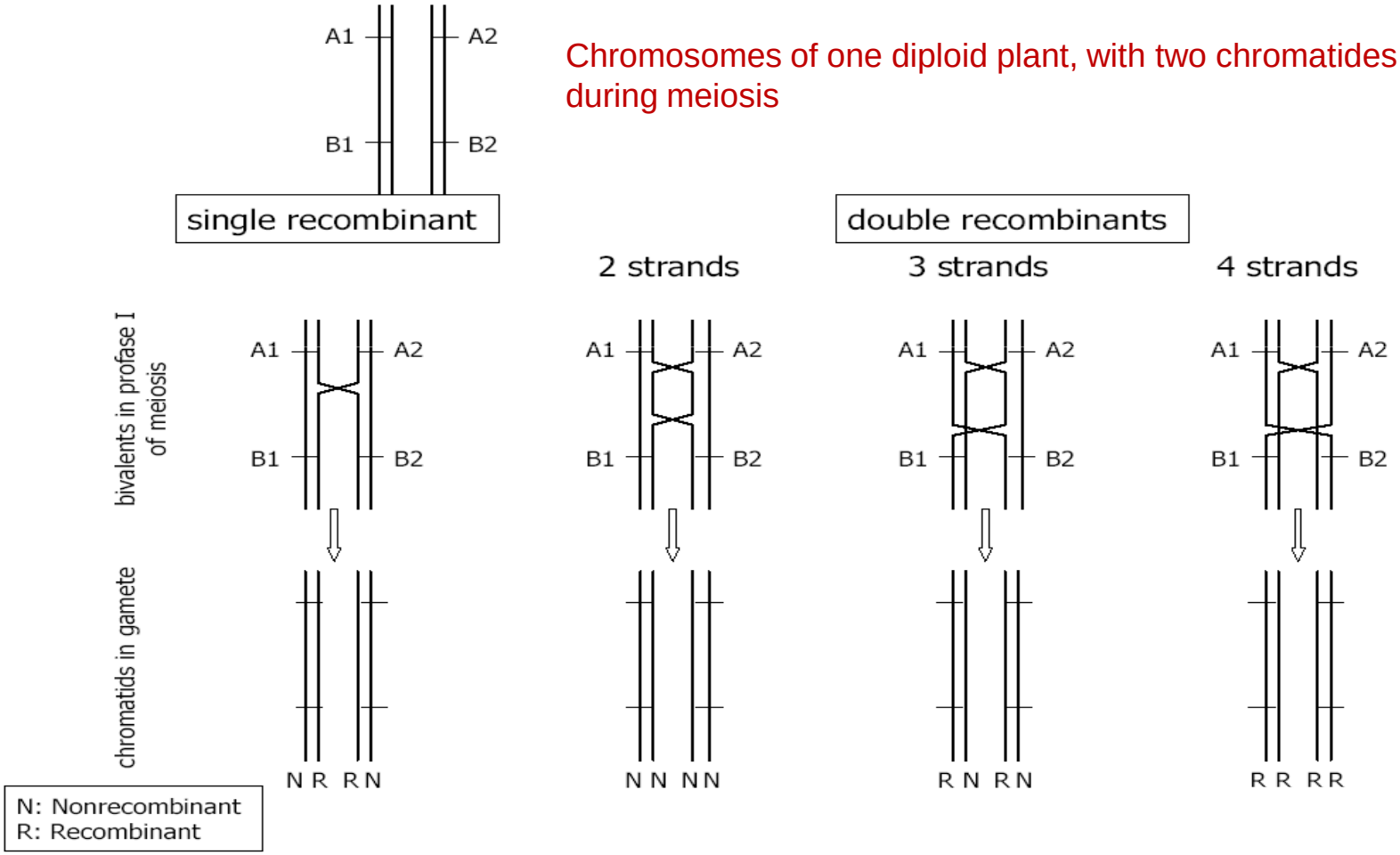
- Involves the exchange of pieces between homologous chromosomes
- It occurs during the meiosis I, when the two homologous chromosomes are aligned



Some basic concepts

LINKAGE AND RECOMBINATION

Chromosomes of one diploid plant, with two chromatides each, during meiosis



Some basic concepts

CONSEQUENCES

- Alleles of loci located in the same chromosome tend to be inherited together, and will not assort independently
- The closer the loci are the higher the chance that their alleles will be transmitted together to the progeny
- The chance of recombination is not homogeneous along the chromosomes; in some regions it happens more frequently than in other regions (e.g. recombination is suppression around the centromeres)
- After a number of generations, numerous ‘new’ combinations of alleles at different loci are formed in the population

Some basic concepts

LINKAGE VS. ASSOCIATION

- Association is a statistical statement about the co-occurrence of alleles and/or phenotypes:
  - Plants with resistance **R** have allele **A** more often than expected => allele **A** is associated with resistance **R**
- Linkage is a genetic relationship between loci; linkage does not of itself produce any association in the general population
  - Linkage creates associations (linkage disequilibrium) within families, but this does not necessarily hold in the general population

However:

- If two 'unrelated' plants with resistance **R** have inherited it from a common ancestor, they might also tend to share particular ancestral alleles at loci linked to **R** (linkage disequilibrium is still present at the level of the population of unrelated individuals)

ILVO

Some basic concepts

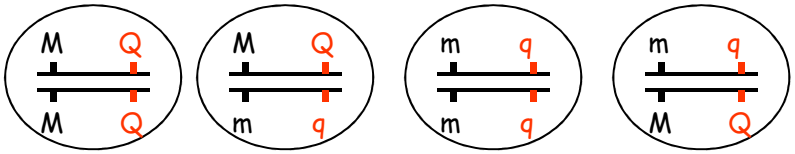
LINKAGE VS. ASSOCIATION

- Statistical association
  - In families
    - Is due to linkage (linkage disequilibrium exists between the alleles at the linked loci)
  - In ‘unrelated’ individuals
    - Can be due to linkage (linkage disequilibrium exists between the alleles at the linked loci)
    - Can be due to other causes (population stratification, selection, admixture of populations, …)

Some basic concepts

LINKAGE VS. ASSOCIATION

- Markers linked to a gene are in LD with the gene (= some marker-gene combinations are more frequent than expected)



	exp	obs
MQ	2	4
Mq	2	0
mQ	2	0
mq	2	4

- LD is created artificially in segregating populations useful for linkage map construction
- LD is also present in natural and breeding populations => the principle can be used to identify markers closely linked to genes (or QTL) of interest using non-structured populations

DNA-polymorphisms
DNA-markers

Polymorphisms and Markers

DNA-POLYMORPHISM

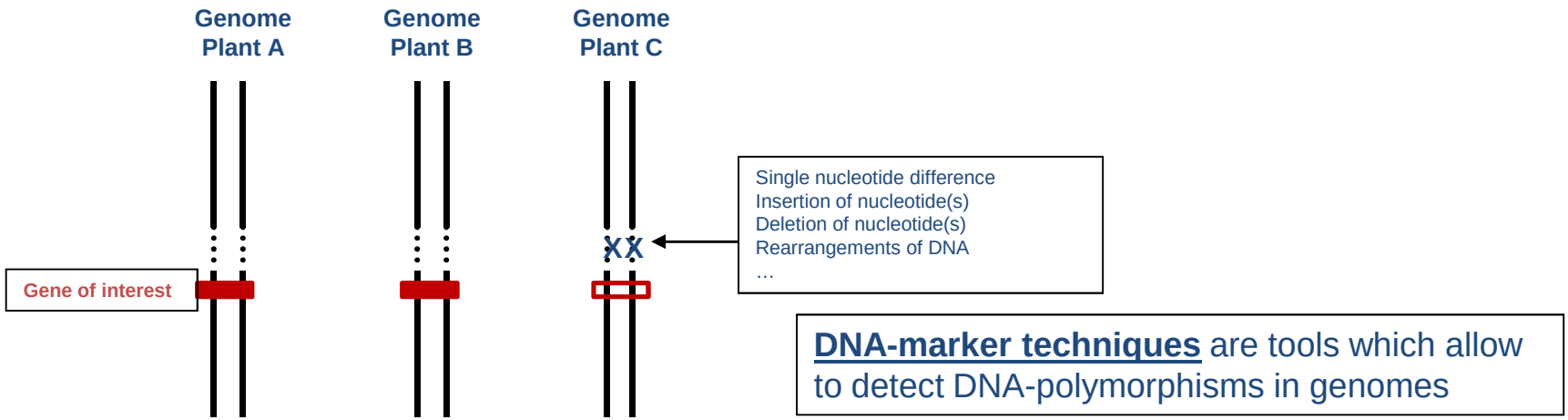
- Definition: “One of two or more alternate forms (alleles) of a chromosomal locus that differ in nucleotide sequence or have variable numbers of repeated nucleotide units” (http://www.biochem.northwestern.edu/holmgren/Glossary/index.html)
  - To know whether a chromosomal locus is polymorphic, we need to compare genomes (for example, the two haploid genomes of a diploid organism or the two diploid genomes of two diploid organisms)
  - DNA-polymorphisms can be found in coding and in non-coding regions
- DNA-polymorphism within organisms:
  - Heterozygous locus: the two alleles present at a given locus in a diploid organism differ
  - Homozygous locus: the two alleles present at a given locus in a diploid organism do not differ
  - Imagine two linked loci A and B (located next to each other); locus A is heterozygous; can you conclude that locus B is also heterozygous?

ILVO

Polymorphisms and Markers

DNA-MARKER

- A DNA-marker is a sequence of DNA, found at a specific location of the genome, which can be readily detected and whose inheritance can be monitored
  - It is the variation in, or the polymorphism of, molecular markers which can be used
  - The term ‘DNA-marker’ is not synonymous of ‘gene’ as most DNA-markers are located in non-coding regions and they do not have any biological function



DNA-marker techniques are tools which allow to detect DNA-polymorphisms in genomes


Samples Plot
File Edit View Tools Alleles Help
Plot Setting: Microsatellite Default
AFLP HOSSEIN E-ACT+
AFLP HOSSEIN E-
Samples Plot
AUTO/EXIT
POWER

DNA-markers

DNA-Marker

DEFINITION

A DNA-marker technique is a procedure to detect DNA-polymorphisms between chromosomes:

- within the same individual (homozygous / heterozygous)
- between individuals

The purpose is to reveal polymorphisms between chromosomes without actually sequencing the DNA

Different marker techniques (procedures) differ in:

- Protocol
- Kind of data obtained (information content, resolution…)
- Number of loci (markers) which are analyzed simultaneously
- Field of application
- Technical and financial requirements

DNA-Marker

INHERITANCE

- The transmission of DNA-markers follows Mendel's laws
- During gamete formation the segregation of the alleles of one allelic pair (=locus) is usually independent of the segregation of the alleles of another allelic pair (=locus)
- Different DNA-marker 'forms' at a given locus are also called alleles
- A diploid plant posses two DNA-marker alleles at a given locus
  - There exist homozygous (2x the same allele) and heterozygous (2 different alleles) plants
  - A plant can be homozygous for one locus and heterozygous for another locus in the neighborhood
- For nuclear DNA-markers: one DNA-marker allele was inherited from the father and another allele was inherited from the mother
- There exist dominant, co-dominant and recessive DNA-marker alleles. Difference with respect to dominance/co-dominance relationships in genes:
  - In the case of genes the classification makes reference to the contribution of the alleles to the phenotype of the organism
  - In the case of DNA-markers this classification makes reference to the relationship to the number of alleles that can be visualized by a given detection system
- If one trait is influenced by more than one gene, DNA-markers in different genomic loci will show association with the trait

ILVO

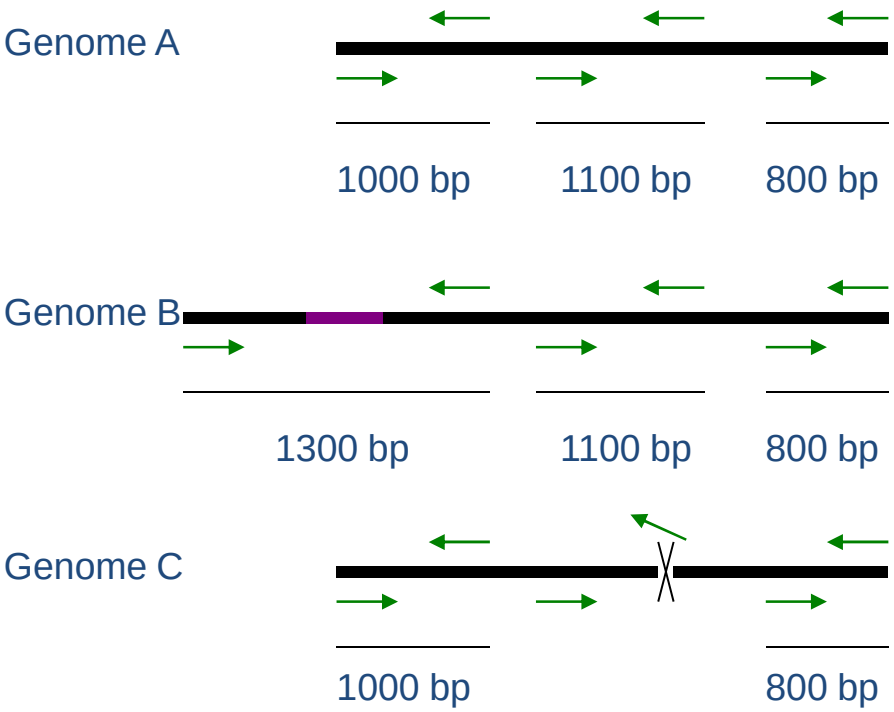
DNA-Marker

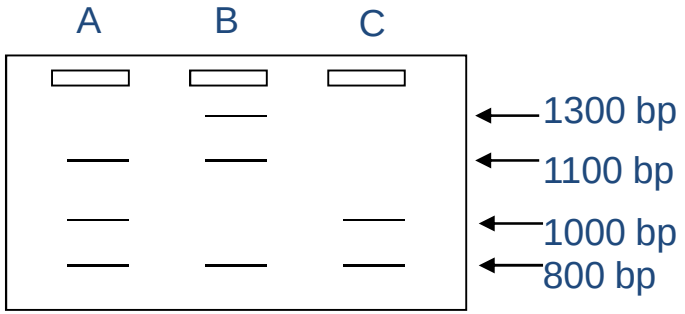
RAPD: DEFINITION

- Random Amplified Polymorphic DNA (Williams et al. 1990)
- Involves the PCR amplification of DNA fragments using short primers (+/- 10 nucleotides)
- Amplification occurs only where sequences complementary to the primers are in close enough proximity for successful PCR
- PCR based => requires little quantities of DNA
- Technically simple
- Disadvantage: reproducibility can be low. DNA template quality, PCR conditions, reagents and equipment can affect the outcome
- Still used, but some journals do not publish RAPD-based studies

DNA-Marker

RAPD: THE TECHNIQUE



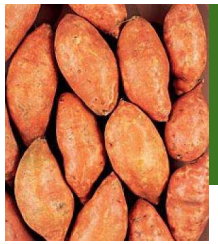


DNA-Marker

RAPD: STEPS

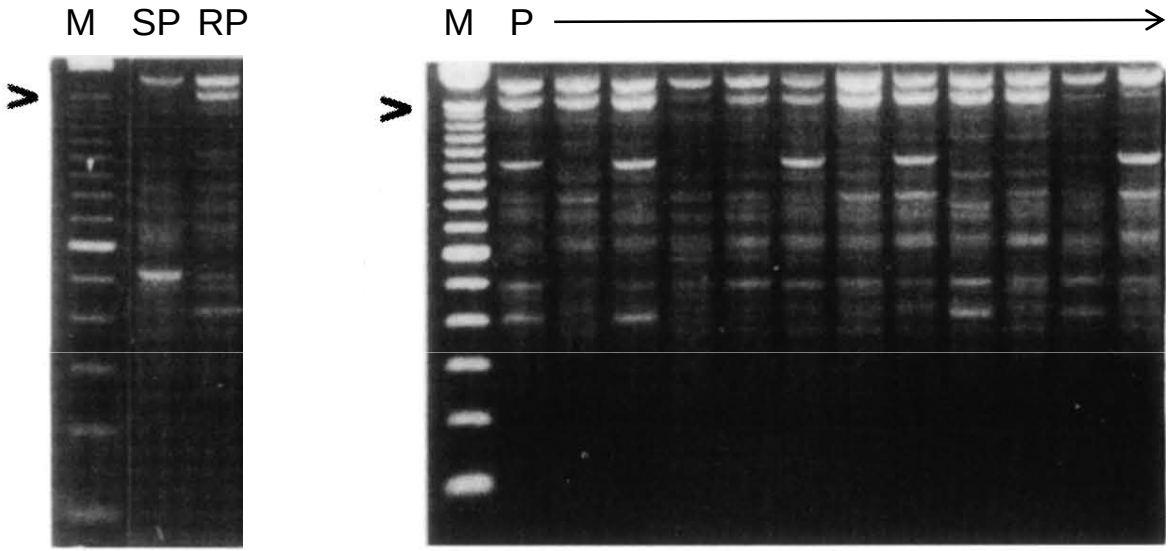
1. Extract DNA
2. Amplify DNA fragments in a PCR reaction using one short primer
3. Separate the PCR products by agarose gel electrophoresis
4. ‘Visualize’ the DNA-fragments (e.g. ethidium bromide)

Usually multiple loci are amplified and differences in band pattern reflect genetic differences



DNA-Marker

RAPD: ROOT-KNOT NEMATODE RESISTANCE IN SWEETPOTATO

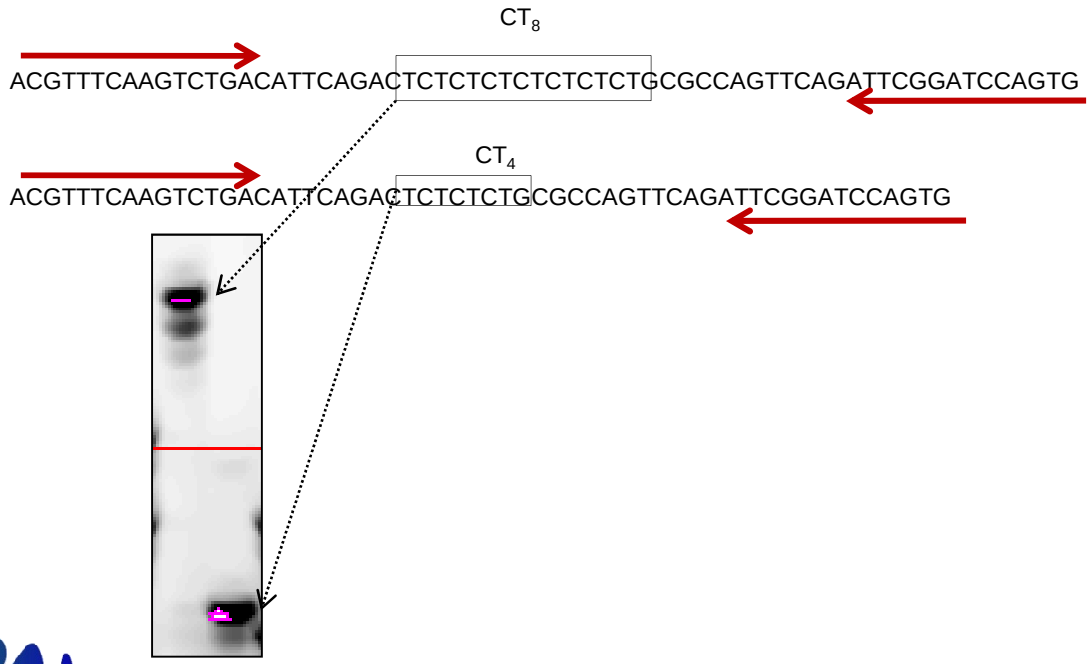


- Ukoskit et al (1997): marker linked to root-knot nematode resistance gene
- RAPD fingerprints of
  - RP: resistant parent
  - SP: susceptible parent
  - P: resistant progeny plants
  - M: length marker

DNA-Marker

SSR: DEFINITION

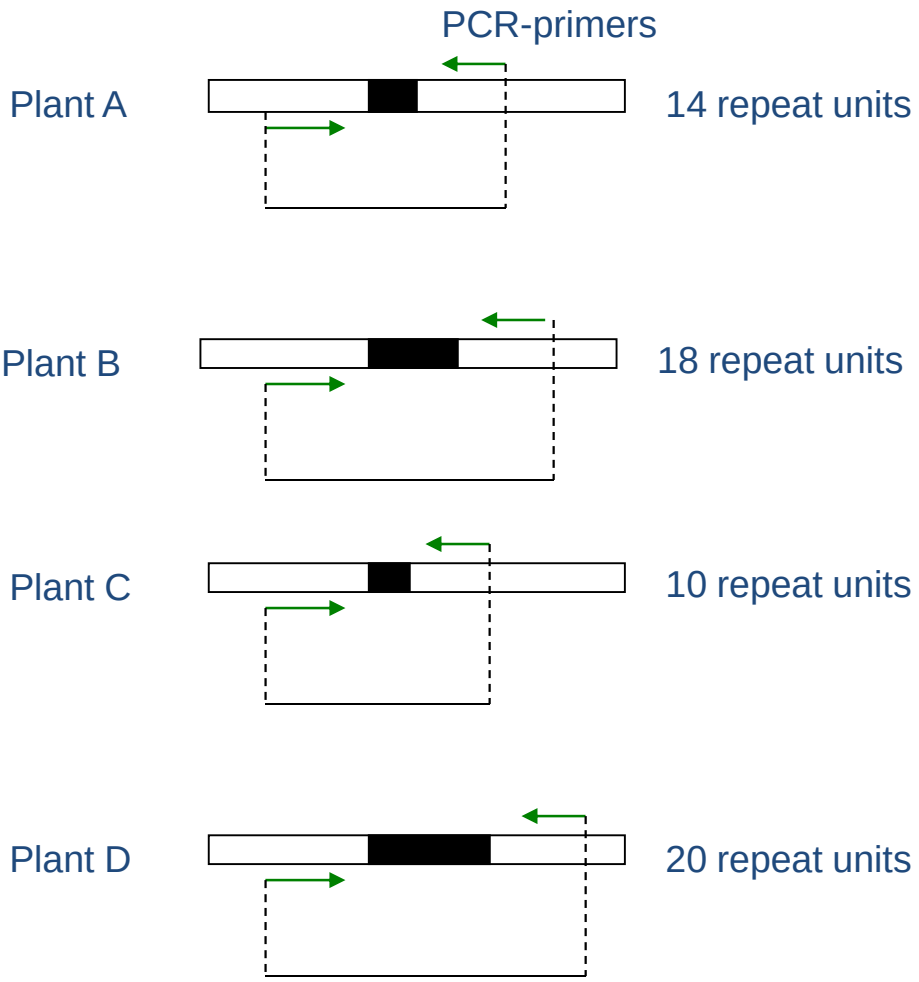
- Simple Sequence Repeat (Grist et al. 1993): involves the study of the variation in number of repeat units at a given locus
  - Repeat units are short (e.g. di-nucleotide repeats, tri-nucleotide repeats)
  - Repeats can be simple [ACACACAC $(AC)_4$; ACGACGACGACG $(ACG)_4$] or interrupted [ACACACACTGTACACAC; $(AC)_4TGT(AC)_3$]
- Examines differences in size of PCR-amplified DNA fragments
- PCR-based => requires little amounts of DNA

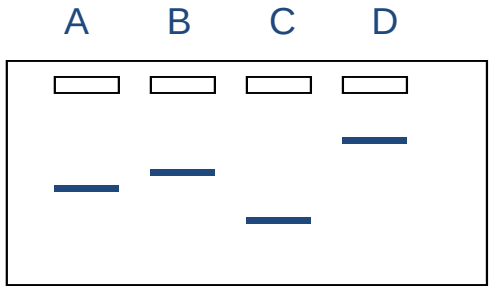


Primer sequences are designed in conserved regions

DNA-Marker

SSR: THE TECHNIQUE





Note: all individuals in the example are homozygous for this locus

Potential problems:

1. size homoplasy (false homology) of SSR alleles,
2. allele reversion

DNA-Marker

SSR: STEPS

1. Extract DNA
2. Amplify DNA fragments in a PCR reaction using two highly specific primers
   Finding these primers can be time-consuming and expensive. Several approaches can be followed
3. Separate the fragments by (polyacrylamide) gel electrophoresis
4. ‘Visualize’ the amplified DNA-fragments (e.g. ethidium bromide / silver staining / fluorescent labeling of primers)

Differences in band length reflect genetic differences (between chromosomes of the same individual or between chromosomes of different individuals)

DNA-Marker

SSR: APPLE DIVERSITY

- SSR fingerprints of 4 apple genotypes using Applied Biosystems capillary technology (AB3130 sequencer)
- For each locus 1 (homozygote) or 2 (heterozygote) bands are amplified
- Up to 4 colours can be analysed simultaneously

DNA-Marker

SSR: PRIMER PAIRS

- Primers are derived from DNA sequence information in the regions flanking the SSR-motif

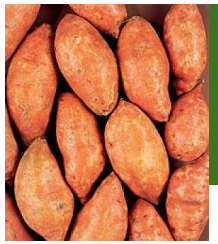
- This information is derived from:
  - Knowledge in other (related) species
    - Some primer pairs *I. trifida* can be used in *I. batatas, I. tiliacea, I. triloba and I. lacunosa* - Hu et al (2004)
  - ~~Genomic libraries / cDNA libraries~~
    - ~~Development of library~~
    - ~~Enrichment for SSR sequences (optional)~~
    - ~~Identification of clones containing SSR inserts~~
    - ~~Sequencing of these clones~~
    - ~~Primer pair development~~
  - Databases (search for SSR motifs in EST sequences) of different origin) - Wang et al 2011…
  - Transcriptome sequencing using Next Generation technologies (Wang et al 2010)

Table 4
Allele sizes (bp) of database-derived-SSRs for *Ipomoea* species. Polymorphisms (P) and transferability (Trans) between genotypes (G) and species (S) are shown

SSR	*I. trifida*	*I. batatas*	*I. lacunosa*	*I. triloba*	*I. tiliacea*	P	Trans	T_a (°C)
ITSSR01	222, 224, 230	216, 222, 226, 232,	228	222	224	G, S	A	64
ITSSR02	158, 162	167, 192, 194, 198, 200	180	174	167	G, S	A	64
ITSSR03	142, 162, 178	142, 162, 164, 170	n.a.	n.a.	n.a.	G, S	S	60
ITSSR04	266, 272, 276	264, 266, 268, 272	268	268	268	G, S	S	64
ITSSR06	177	177, 179, 183	n.a.	n.a.	181	G, S	S	50
ITSSR07	168, 171	171, 174	171	168	171	G, S	S	57
ITSSR08	175, 181	175, 181	175	175	178	G, S	S	57
ITSSR09	135	111, 135, 156	n.a.	188	129	G, S	S	58–53
ITSSR11	118, 122	122	122	122	118	G, S	S	65
ITSSR12	122, 117	127, 122	117	122	122	G, S	A	64
ITSSR14	175, 163	163	193	151	157	G, S	A	55
ITSSR15	132	132, 136	n.a.	n.a.	n.a.	S	S	64

A: transferable to all accessions in this study; n.a.: no amplification; T_a: annealing temperature.

12 SSR loci



DNA-Marker

SSR: SWEETPOTATO PRIMER PAIRS

Wang et al (2011) – EST databases

~300 - 1400 SSR loci

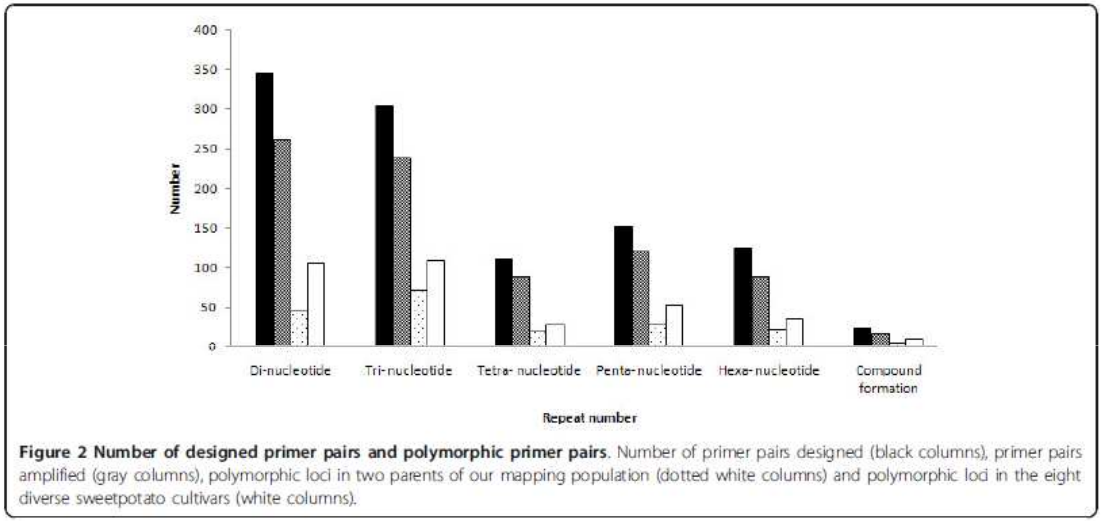


Figure 2 Number of designed primer pairs and polymorphic primer pairs. Number of primer pairs designed (black columns), primer pairs amplified (gray columns), polymorphic loci in two parents of our mapping population (dotted white columns) and polymorphic loci in the eight diverse sweetpotato cultivars (white columns).

Wang et al (2010): Illumina RNA-seq data

~4000 SSR loci

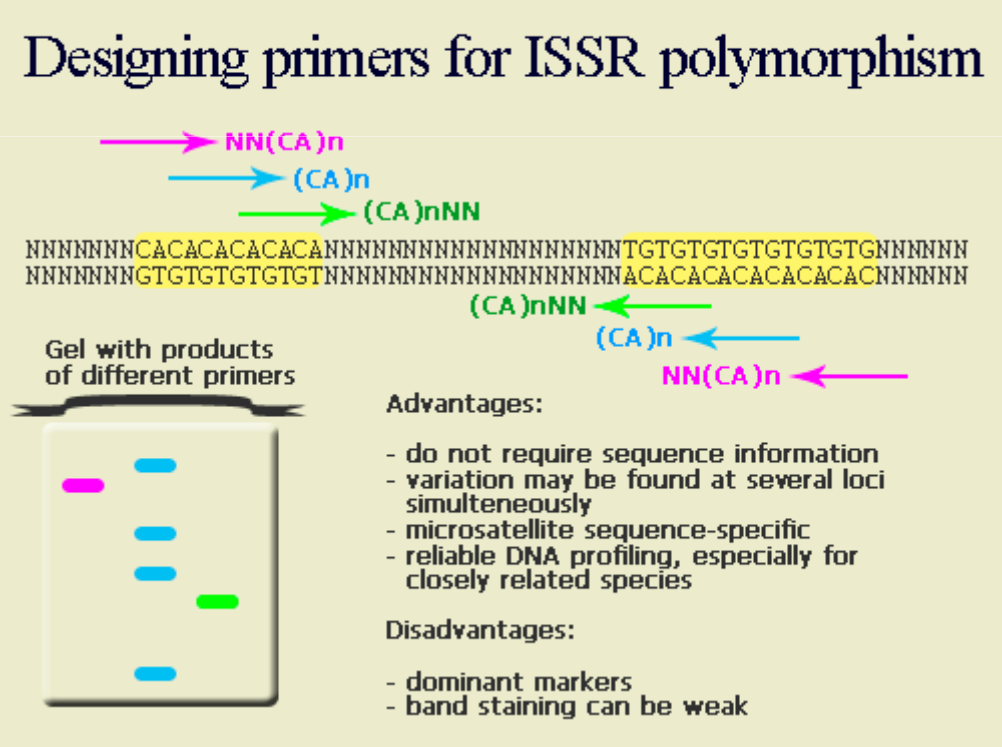
Table 2 Summary of cSSR searching results

Searching Item	Numbers
Total number of sequences examined	56,516
Total size of examined sequences (bp)	32,852,951
Total number of identified cSSRs	4,114
Number of cSSR containing sequences	3,594
Number of sequences containing more than 1 cSSR	423
Number of cSSRs present in compound formation	275
Di-nucleotide	1782
Tri-nucleotide	1747
Tetra-nucleotide	330
Penta-nucleotide	142
Hexa-nucleotide	113

DNA-Marker

ISSR: DEFINITION

- Lenght polymorphisms found between microsatellite repeats
- Primers are designed based on microsatellite repeats and target multiple loci (eventually extended with selective nucleotides)

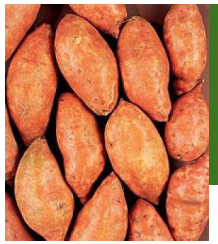


DNA-Marker

ISSR: STEPS

1. Extract DNA
2. Amplify DNA fragments in a PCR reaction using two repeat-containing primers
3. Separate the fragments by (polyacrylamide) gel electrophoresis
4. ‘Visualize’ the amplified DNA-fragments (e.g. ethidium bromide / silver staining / fluorescent labeling of primers)

Differences in band length reflect genetic differences (between chromosomes of the same individual or between chromosomes of different individuals)



DNA-Marker

ISSR: SWEETPOTATO ACCESSIONS

- Huang & Sun (2000):
  - Comparison of accessions of *Ipomea* series *Batatas*

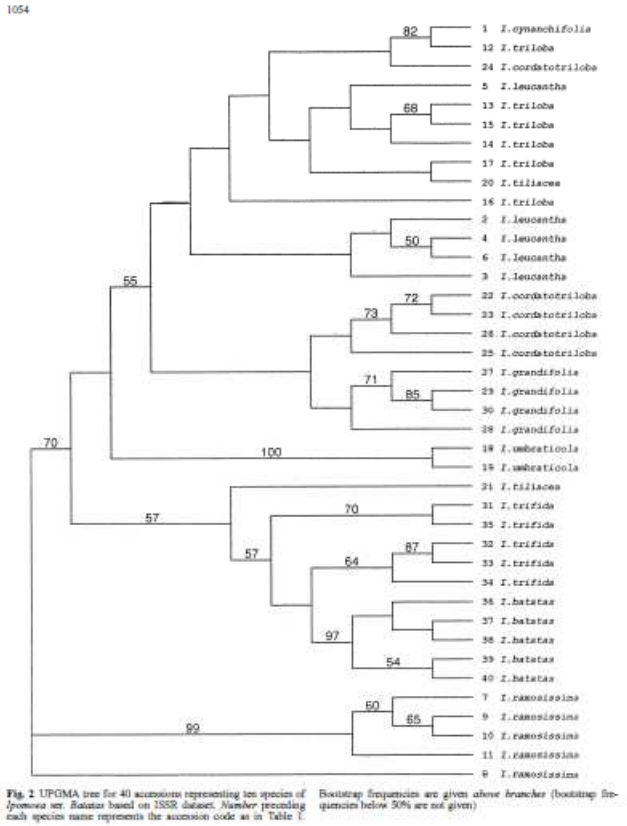
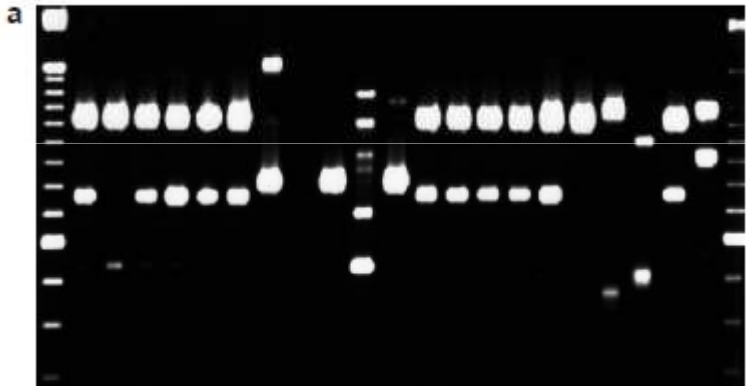


Fig. 2 UPGMA tree for 40 accessions representing ten species of *Ipomoea* ser. *Batatas* based on ISSR dataset. *Number* preceding each species name represents the accession code as in Table 1. Bootstrap frequencies are given *above branches* (bootstrap frequencies below 50% are not given)



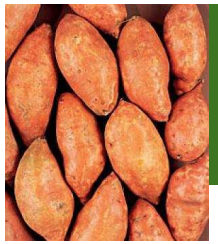
DNA-Marker

CAPS

- Cleaved Amplified Polymorphic Sequences
- Restriction site polymorphisms in PCR products

1. Extract DNA
2. Amplify DNA fragments in a PCR reaction using two repeat-containing primers
3. Separate the fragments by (polyacrylamide) gel electrophoresis
4. ‘Visualize’ the amplified DNA-fragments (e.g. ethidium bromide / silver staining / fluorescent labeling of primers)

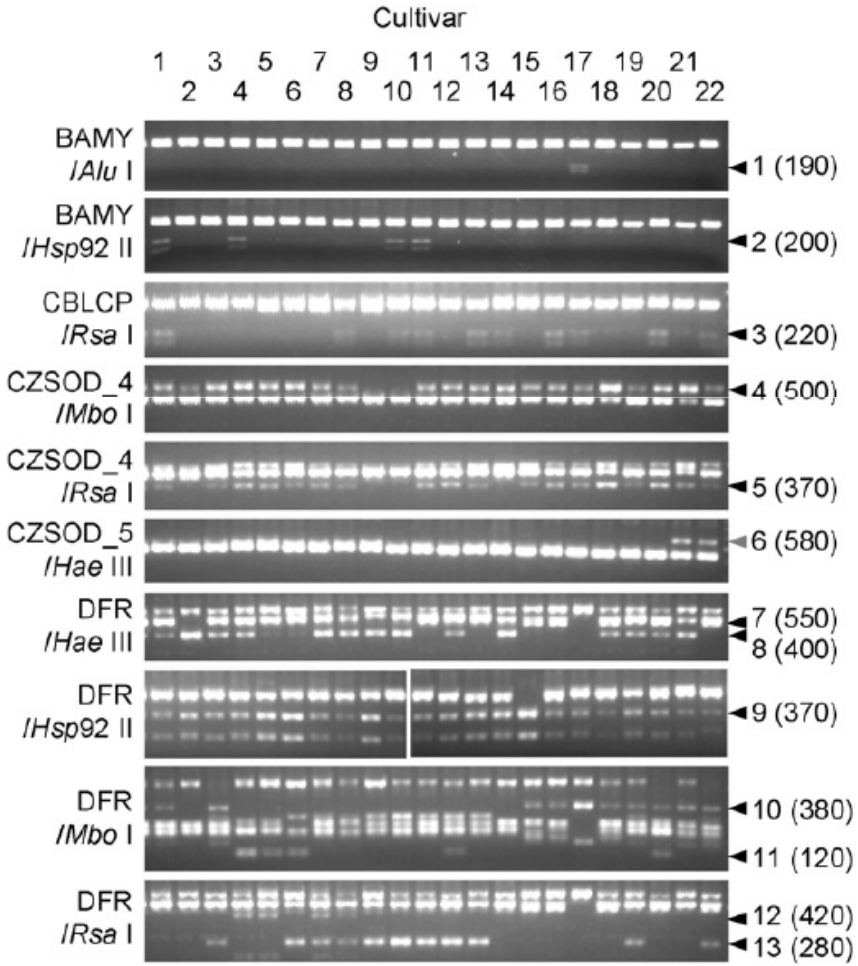
Differences in band length reflect genetic differences (between chromosomes of the same individual or between chromosomes of different individuals)



DNA-Marker

CAPS: SWEETPOTATO CULTIVARS

- Tanaka et al (2010):
  - ✓ Comparison of sweetpotato cultivars
  - ✓ Arrowheads are polymorphic markers
  - ✓ BAMY= Beta-amylase
  - ✓ CBLCP= Cathepsin B-like cysteine proteinase
  - ✓ Alu I, Hsp92, etc are restriction enzymes



DNA-Marker

AFLP: DEFINITION

- Amplified Fragment Length Polymorphism (Vos et al. 1995): identifies Lenght polymorphisms in amplified restriction fragments
- Examines multiple loci in one single reaction => high multiplex ratio
- Requires little amounts of DNA
- Fluorescent labeling and detection on automatic sequencers (LiCor, AB), allow some automation
- Several steps involved => relatively difficult to perform, skilled operators required

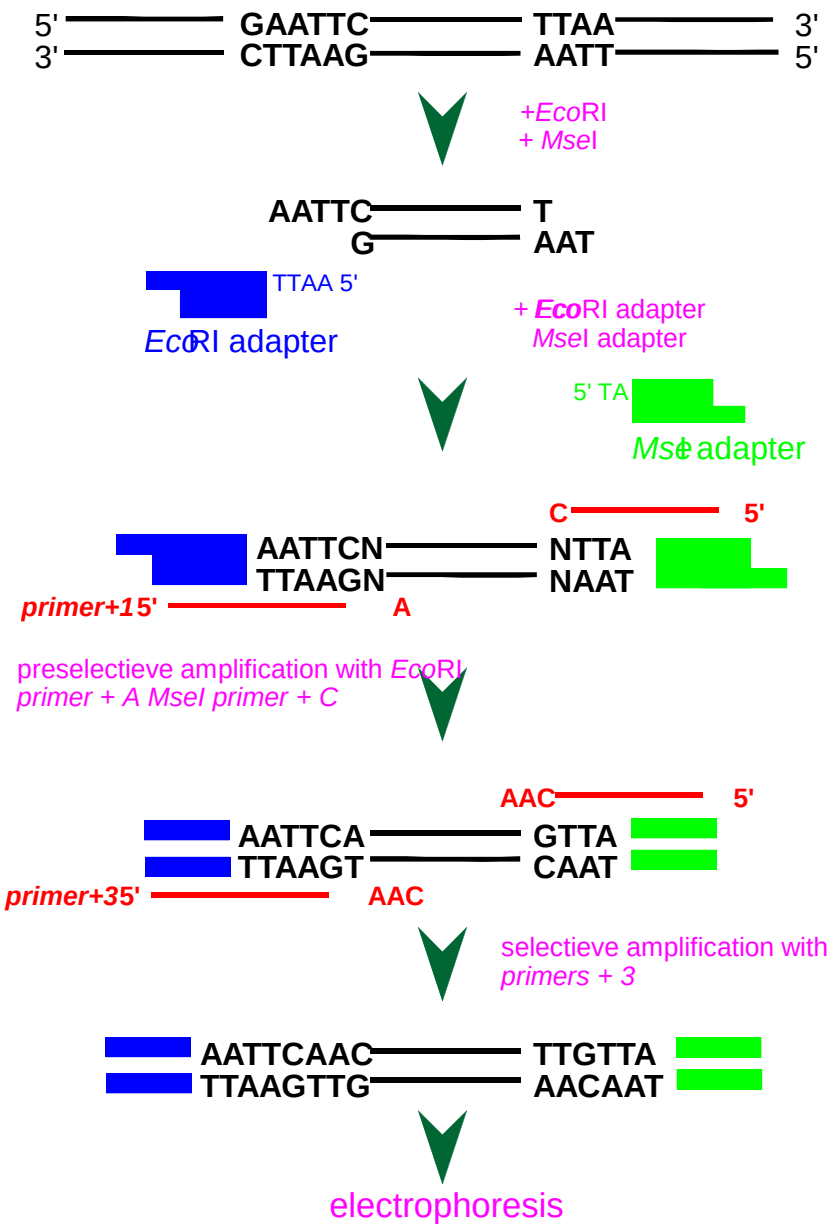
DNA-Marker

AFLP: STEPS

1. Extract DNA
2. Digest the DNA using two restriction enzymes. One Hexacutter (e.g. *Eco*RI) and one Tetracutter (e.g. *Mse*I)
3. Ligate double stranded adaptors to the restriction fragments
4. Perform one PCR amplification using primers complementary to the adaptor sequence.
   - These primers are usually extended with 1 or 2 ‘selective nucleotides’
5. Perform a second PCR amplification using labeled primers complementary to the adaptor sequence
   - These primers are usually extended with 1-4 ‘selective nucleotides’

The total number of selective nucleotides required depends on the complexity of the genome to be analyzed:

Bacteria: 0-2 selective nucleotides
Fungi: 3-5 selective nucleotides
Plants: 5-8 selective nucleotides

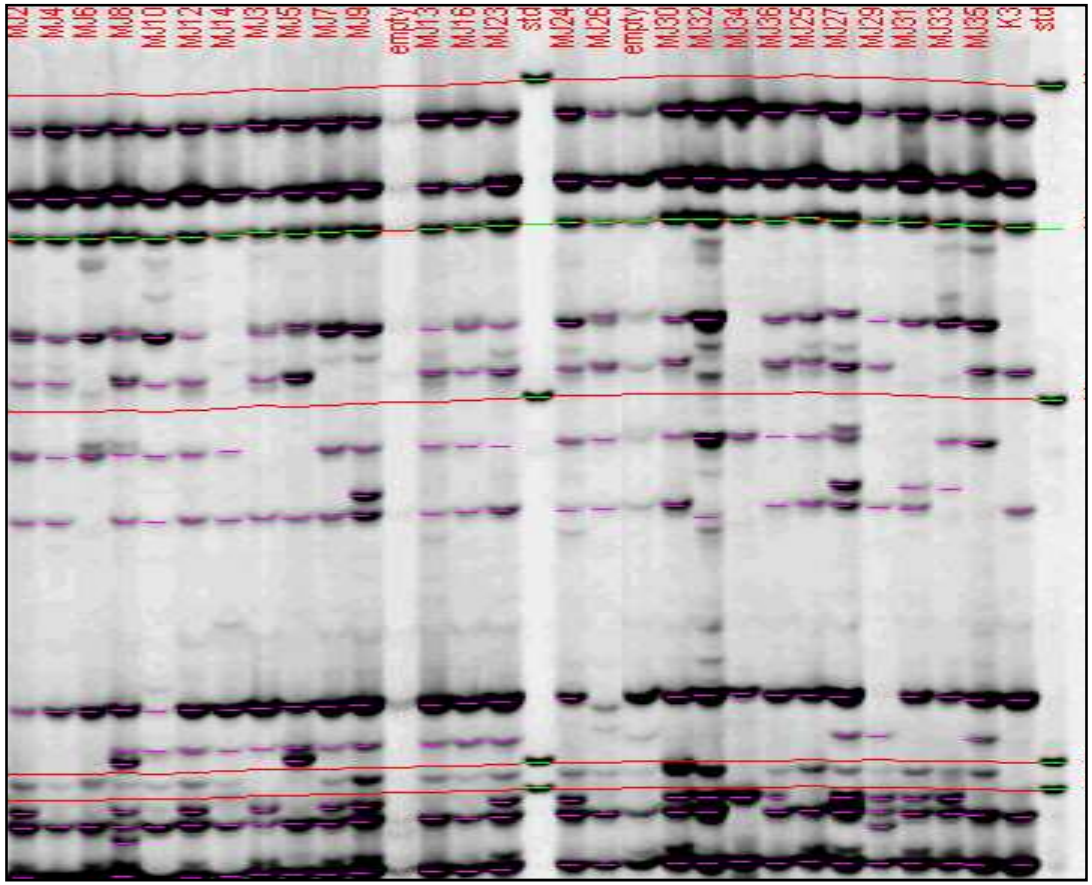


- Basis of the polymorphisms detected:
  - Changes in the restriction enzymes recognition sites
  - Changes in the recognition sites of the primer extensions (selective nucleotides)
  - Insertions/deletions within the DNA fragment
- Multilocus fingerprint is generated. In most cases, impossible to determine allelic relationships between AFLP fragments

Note: each addition of one selective nucleotide will result in a reduction of ¼ of fragments amplified
=> This is one of the most important characteristics of the AFLP technique: 'flexibility'

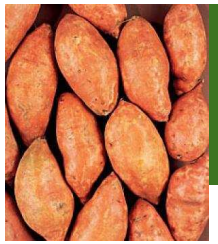
DNA-Marker

AFLP: *M. SYLVESTRIS* GENOTYPES FOM FLANDERS



- Each lane is the DNA-fingerpring of a different plant
- Each band within a lane represents a DNA-fragment; different bands within a lane represent different genetic loci
- Bands from different lanes that migrate to the same distance represent the same genetic locus in different individuals

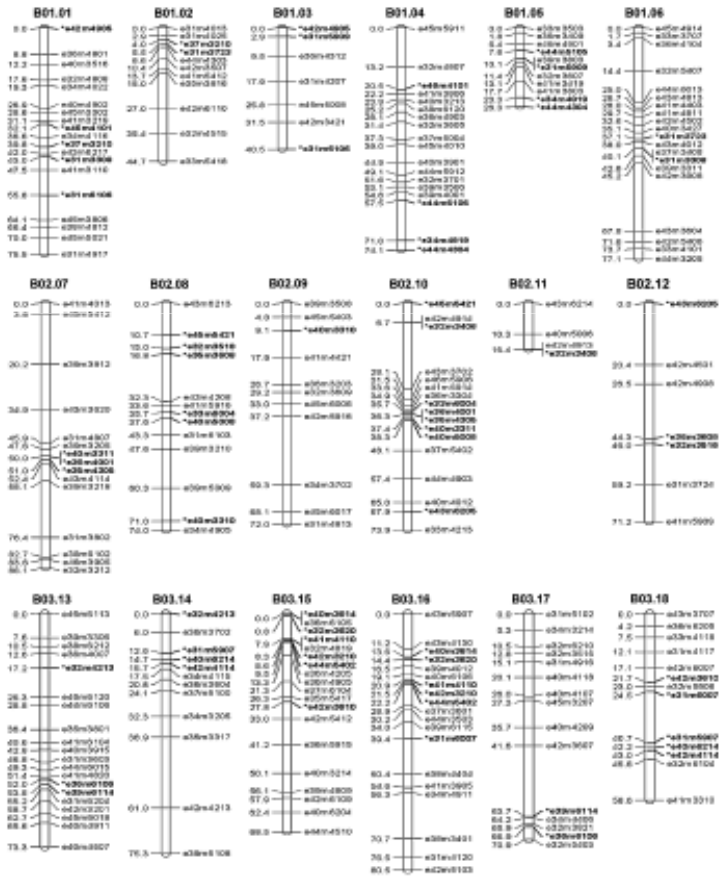
Std: size standard



DNA-Marker

AFLP: GENETIC MAP OF SWEETPOTATO

- Cervantes-Flores et al (2008)



DNA-Marker

SPN: DEFINITION

- Direct analysis of genetic variation at the DNA sequence level: a SNP (Single Nucleotide Polymorphism) is defined as **a single nucleotide position in a given DNA stretch at which there are variations between different individuals of the same species** => formally INDELS (small insertions and deletions) are not SNPs, but this depends on the author

- SNPs are the most common form of DNA sequence polymorphism (more than SSRs and AFLPs) and very abundant in plant genomes
  - ⇒ offer the opportunity to uncover allelic variation directly within expressed sequences of candidate genes (genes 'suspected' to be involved in the trait). Example: allelic diversity at the maize Dwarf8 gene, indicated association with flowering time (Thornsberry JM et al 2001)
  - ⇒ theoretically allow the construction of very dense genetic maps and a high resolution in whole-genome scans

ILVO

DNA-Marker

SNP: BASICS

- Four different types:
  - One transition C⇔T (G⇔A)
  - Three transversions C⇔A (G⇔T); C⇔G (G⇔C); T⇔A (A⇔T)

- Example
  - *Allele 1:* ATGTGCATTA
  - *Allele 2:* ATGTCCATTA

- In principle SNPs could be bi-, tri-, or tetra-allelic polymorphisms. However, tri-allelic and tetra-allelic SNPs are rare almost to the point of no existence, and so SNPs are mostly referred to as bi-allelic markers

DNA-Marker

SNP: MAIZE EXAMPLE

- Ching A et al (2002): SNP frequency, haplotype structure and linkage disequilibrium at 18 loci in 36 elite maize inbred lines
- Maize is highly polymorphic, but reduced allelic diversity is expected in domestication-related genes (breeders select preferentially one of the alleles => reduced allelic diversity)
- Multiple nucleotide changes and indels of various lengths were identified
  - 1 SNP per 60.8 bp (different in coding and non-coding regions)
  - 1 indel per 126.1 bp (different in coding and non-coding regions)
- This translates into a large number of SNPs (and indels) potentially available for use as genetic markers
  - A higher density of genetic markers can be achieved with SNPs than with other marker techniques
- It is feasible to obtain several SNP markers in the vicinity of each maize gene

DNA-Marker

SNP: MAIZE EXAMPLE

Summary of polymorphism analysis (Table 3 Ching et al, 2002)
BMC Genet. 2002; 3: 19.
Published online 2002 October 7. doi: 10.1186/1471-2156-3-19.

Parameter	Value	Comments
Number of loci screened	18	
Total length of amplicons, bp.	6935	2349 coding, 4586 non-coding
Number of bases of sequence screened, bp.	213999	
Number of all sequence variants (SNPs and Indels)	**169**	**1 per 41 bp**
Number of nucleotide substitutions	114	
Transitions / Transversions ratio	1.53	
Frequency of polymorphic sites per bp	0.0164	1 per 60.8 bp
Frequency of polymorphic sites per bp (coding)	0.0077	1 per 130.5 bp
Frequency of polymorphic sites per bp (non-coding)	0.021	1 per 47.7 bp
Number of Indels	55	
Overall indel frequency	0.0079	1 per 126.1 bp
Frequency of indels per bp (coding)	0.0004	1 per 2349 bp
Frequency of indels per bp (non-coding)	0.0118	1 per 85 bp

DNA-Marker

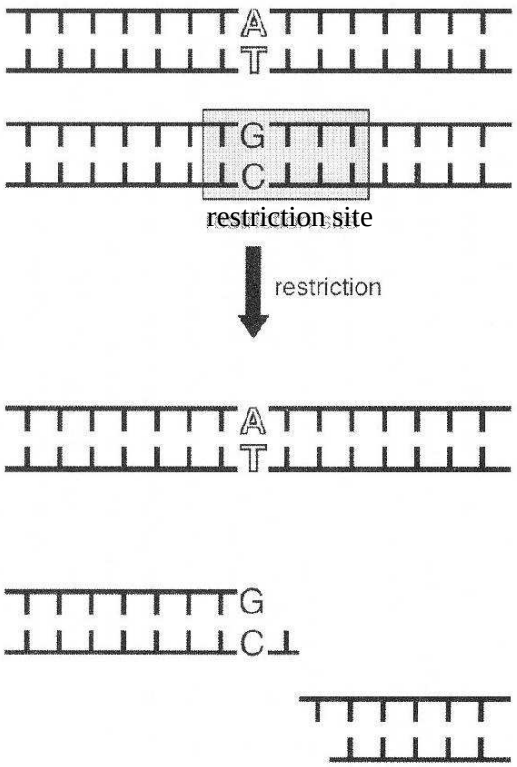
SNP: DETECTION

- Large variety of instrumentation, software, and reagent systems
- Two main types:
  - Abbreviated form of DNA sequencing technology
    - Primer placed immediately upstream of the allele in question and incorporate a single nucleotide at the SNP site => fluorescent-based detection (SNP-IT (BioScience); SNaPshot(Applied Biosystems), or mass determination (MALDI-TOF, Sequenom) can be used to discriminate the SNP allele
    - Primer placed near the SNP site, allowing the sequencing reaction to proceed to, or through, the SNP site (Pyrosequencing)
  - Basic hybridization affinity between perfectly and imperfectly complementary DNA strands as basis for determining the identity of the SNP site
    - CSCE: Conformation Sensitive Capillary Electrophoresis
    - Differential hybridization stringency is used to determine which nucleotide occupies the SNP site

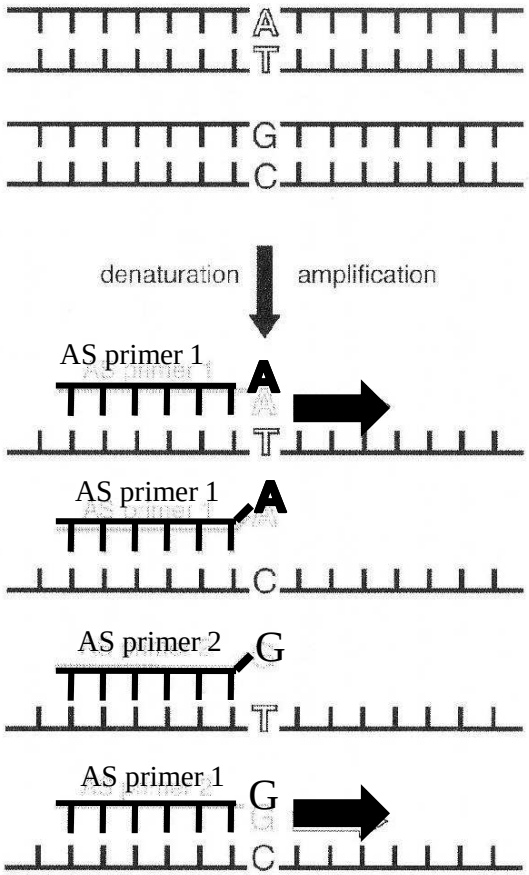
DNA-Marker

SNP: BIOCHEMICAL PRINCIPLES

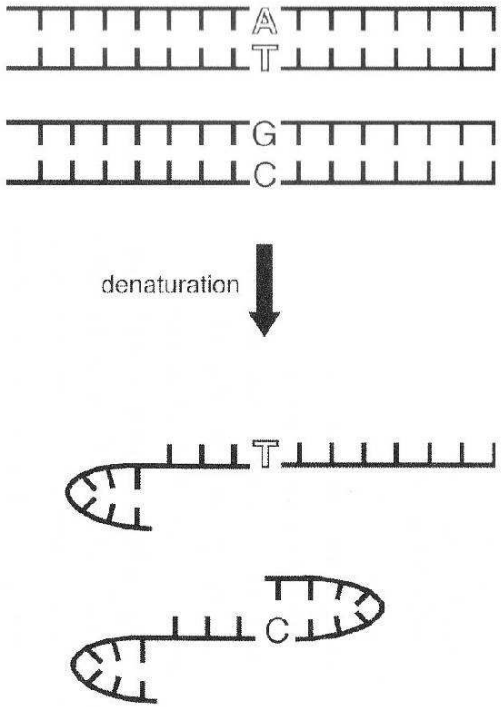
Restriction endonuclease digestion



Allele-specific primer PCR

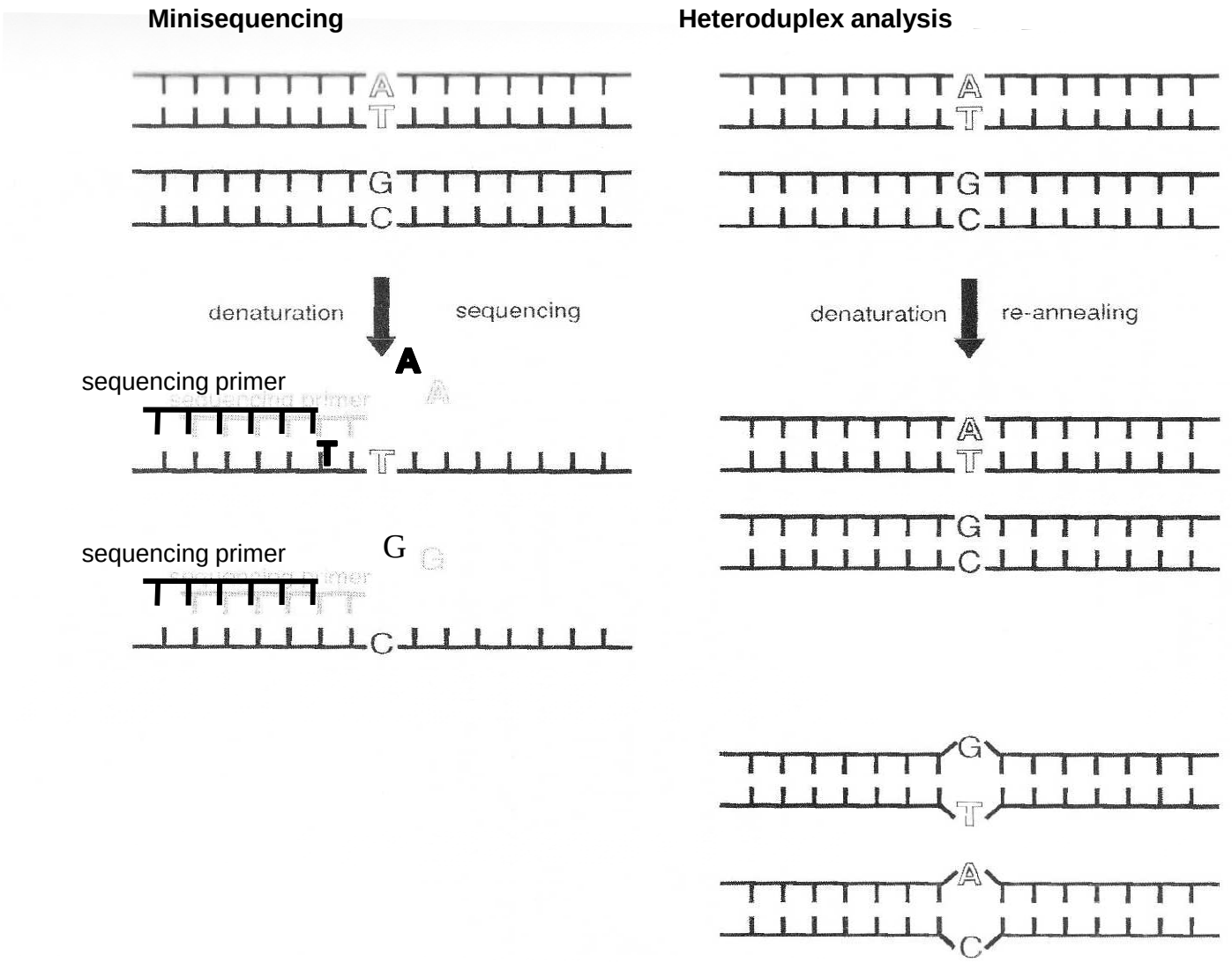


Single-strand conformation polymorphism



DNA-Marker

SNP: BIOCHEMICAL PRINCIPLES



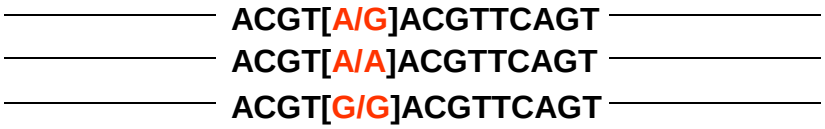
DNA-Marker

SNP: TWO EXAMPLES

- Common steps:
  - Identify target sequence
  - Develop specific primers to PCR-amplify the region of interest
- CSCE: denature and renature the PCR product
  - If locus homozygous, only homoduplices are formed
  - If locus heterozygous, homo- and heteroduplices are formed
  - Heteroduplices and homoduplices have different electrophoretic motilities
- SNaPshot
  - The primer is designed to stop just 5' of the SNP
  - A reaction is performed using ddNTPs labeled with different colours (A= green, C= black, G= blue, T= red)
  - During the reaction the ddNTP corresponding to the allele present is incorporated, and the genotype is read as the peak colors

SNaPshot

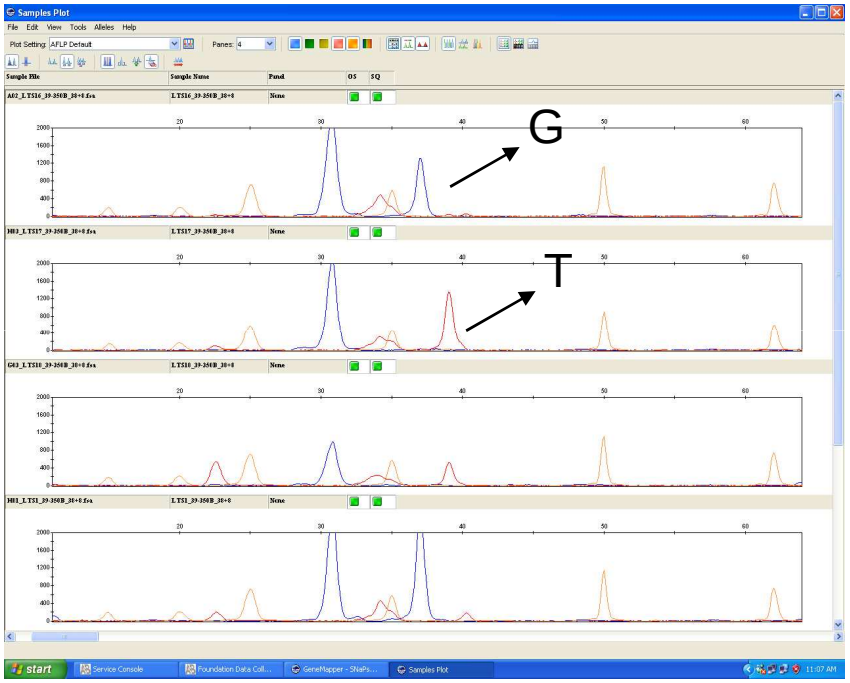
CSCE



DNA-Marker

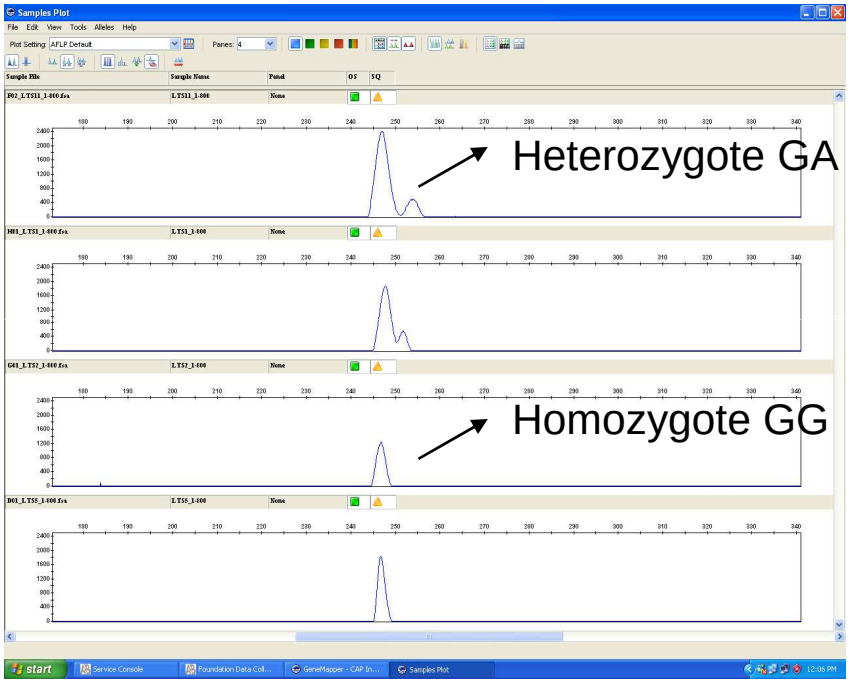
SNP: TWO EXAMPLES

SNaPshot



- Sequencing reaction using dDNTPs, with different fluorescent labels
- We can see which nucleotide is incorporated at the hand of the colour

CSCE



- Makes use of heteroduplex mobility differences
- Heteroduplex has a different migration dynamics than homoduplex

DNA-Markers in sweetpotato

APPLICATIONS

	ISSR	cp-DNA	mt-DNA	AFLP	ITS	SSR	RAPD	CAPS	EST
Interspecific relationships	-Huang & Sun (2000) -Hu *et al* (2003)	-Huang & Sun (2000)	-Huang & Sun (2000)*	-Huang *et al* (2002)	-Huang *et al* (2002)				
Phylogeographical relationships							-Gichuki *et al* (2003)		
Paternity analysis						-Buetler *et al* (2002)			
Linkage mapping	-Hu *et al* (2002)			-Krieger *et al* (2003) -Cervantes-Flores *et al* (2008)			-Usokit *et al* (1997a)		
QTL analysis				-Cervantes-Flores *et al* (2011)					
BSA							-Usokit *et al* (1997b)		
cv identification								-Tanaka *et al* (2010)	
Resource development						-Hu et al (2004) -Schafleitner *et al* (2010) -Wang *et al* (2010) -Wang et al (2011)			-Schafleitner *et al* (2010) -Wang *et al* (2010)

DNA-Marker

TYPES

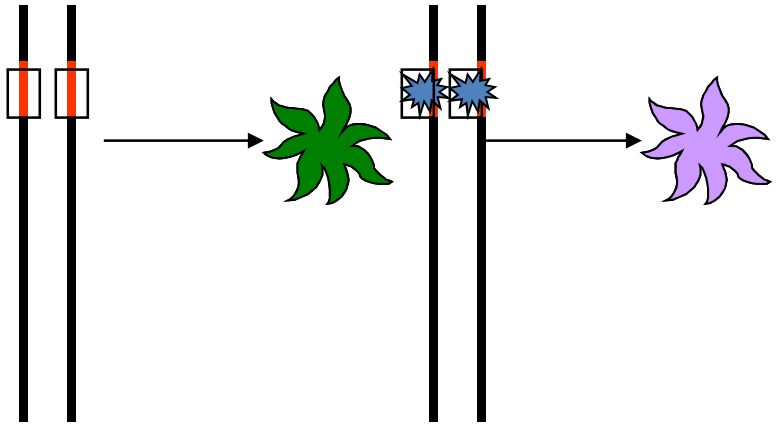
- ‘Function’
  - Functional (detect the mutation which is directly responsible for a change in the phenotype => marker is the ‘causal mutation’)
  - Non functional
    - In the gene
    - In the neighborhood of the gene
- Transmission:
  - Biparental (chromosomal)
  - Uniparental (mitochondria or chloroplast)
- Information-content
  - Co-dominant (visualization of all alleles present)
  - Dominant (visualization of only one allele)

DNA-Marker

FUNCTIONALITY

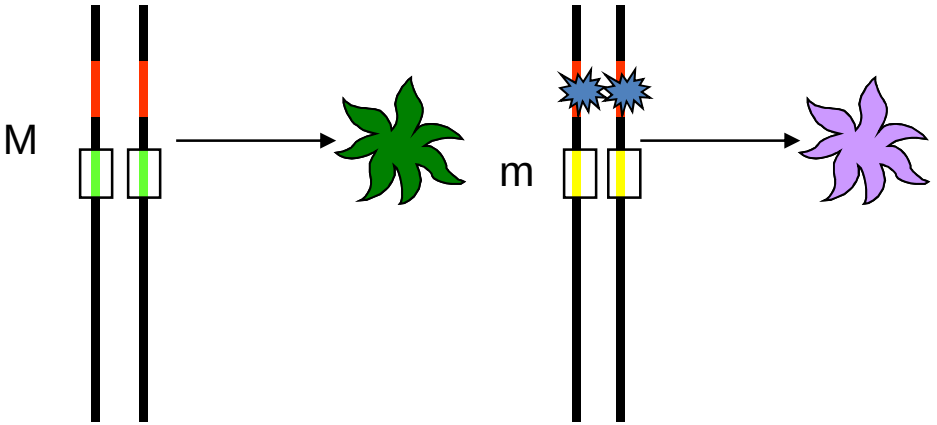
- Functional (detect the mutation which is directly responsible for a change in the phenotype => marker is the 'causal mutation')
- Non functional
  - In the gene
  - In the neighborhood of the gene

Causal mutations: the mutation is responsible for the change in the color of the flower



- The most useful if available
- Difficult to find
- Difficult to prove
- Less often used

Presumed non-functional DNA-markers, in the gene or linked to the gene



- In some cases, enough
- Easier to find
- Easier to prove
- Most frequently used

DNA-Marker

TRANSMISSION

- Uniparental vs. Biparental
  - Biparental nuclear inheritance: nuclear DNA-markers in plants and animals
  - Maternal organellar inheritance: mitochondrial and chloroplast DNA-markers in angiosperms
  - Paternal organellar inheritance: mitochondrial and chloroplast DNA-markers in gymnosperms

DNA-Marker

INFORMATION CONTENT

- Co-dominant:
  - SSR
  - CAPS
  - SNP
- Dominant:
  - ISSR
  - RAPD
  - AFLP