PLNT2530
2024

Unit 6d

Finding a Gene in a Library: Molecular markers

Molecular Biotechnology (Ch 4)

Finding a gene - when you only know the trait

(but not the gene responsible for the trait)

- Disease resistance genes
- Yellow seed coat gene
- Day length insensitivity

Required: The presence/absence of trait must be easily scored

Ideally controlled from a single locus

Strategy: Positional cloning using molecular markers to surround the locus for the genetic trait you seek.

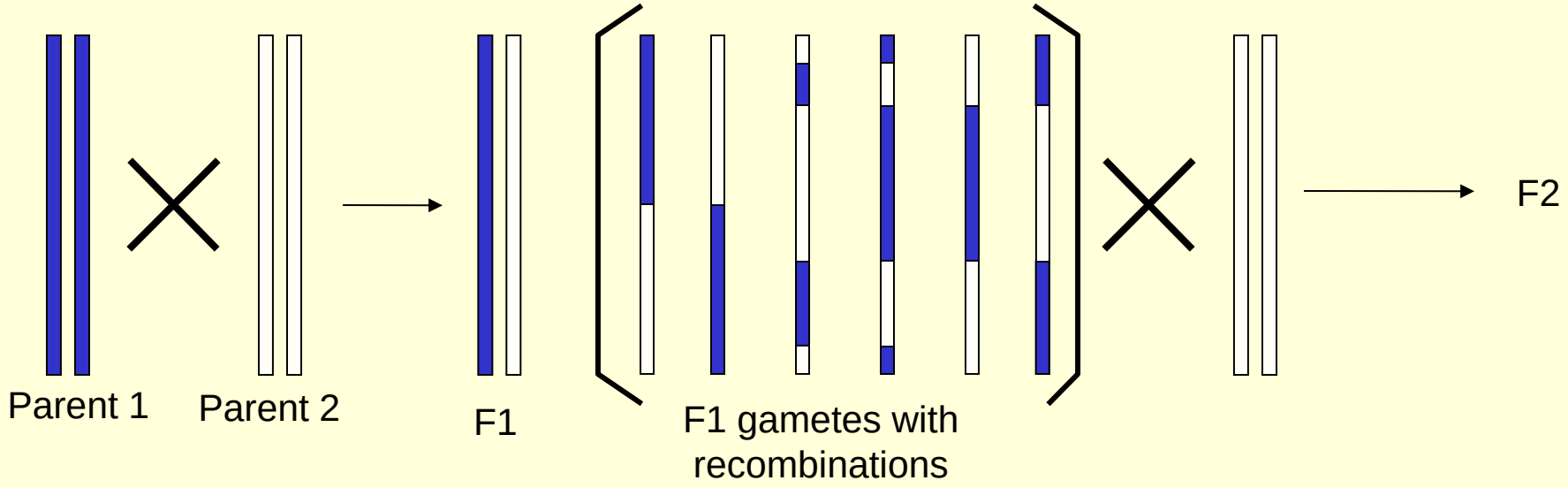
Positional cloning (map based cloning)

- Cloning the gene based on finding the gene by its chromosomal position. Most major crop species already have complete genetic maps with molecular markers covering the entire genome.

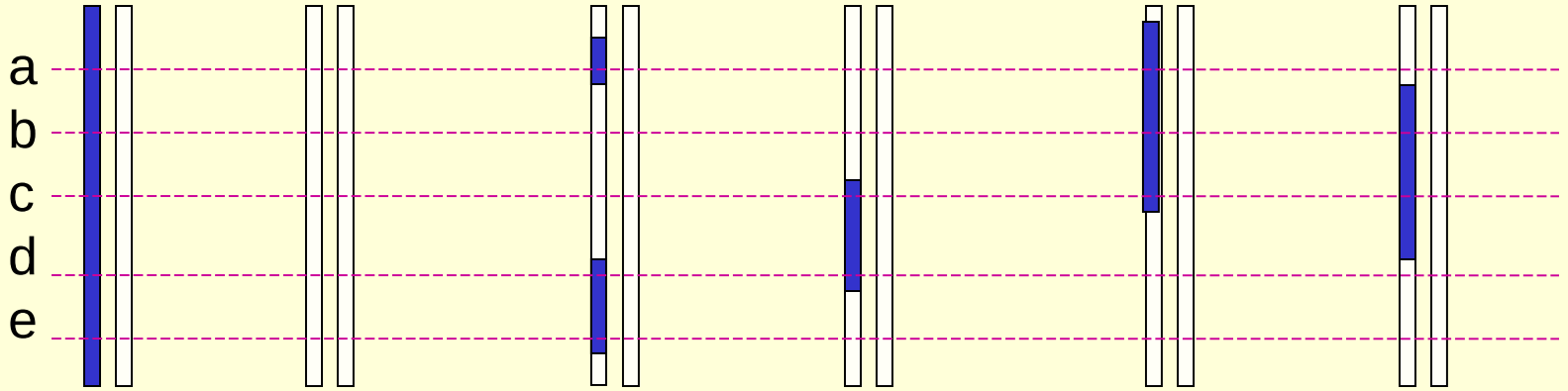
- Molecular markers are probes that can identify genes based on linkage in a genetic cross. Markers act as tags that let us see which parts of a chromosome came from one parent or the other.

- It is extremely unlikely that we'll be lucky enough to find a marker within the gene itself. The goal is to find markers located somewhere "close" to the trait gene, within a few centiMorgans ie. a few percent recombination.

- Requires:
  - Clearly defined, easily measured trait – need to score many plants
  - Homozygous parental plants which <u>do</u> and <u>do not</u> respectively express the trait



F2 progeny



a	+	-	+	-	+	-
b	+	-	-	-	+	+
c	+	-	-	+	+	+
d	+	-	+	+	-	-
e	+	-	+	-	-	-

Linkage analysis – a way of determining the relative arrangement of genes

Genetic method based on recombination frequency for finding how close and in what order genes/markers/traits are to one another.

If 2 trait genes occur close to one another on the same chromosome they will co-segregate with a high frequency

If they lie on different chromosomes they will segregate independently

If they lie on the same chromosome but far apart they will segregate with an intermediate frequency

Positional cloning – an overview

- Obtain a series of molecular markers each representing a unique locus (sequence) on the chromosome – from these identify markers which must be polymorphic for the 2 parent plants (which are polymorphic for your trait)

- Develop a segregating population (F2) from crosses between the 2 parents Compare the segregation of each polymorphic marker site to that of the trait. This shows which are linked and which are not.

- Linkage analysis of the segregating population provides a measure of the distance between the linked molecular markers and the trait gene. This distance is measured in genetic units (centi-Morgan) not physical units (bp) but must be small, 0.5-1 cM (<500,000 bp) for the next step to work.

- Once molecular markers are found close to the trait gene, a process known as chromosome walking is followed by finding contiguous pieces of the chromosome from a genomic library (prepared from the parent possessing the trait) - from the marker site to the trait gene.

- Identify the trait gene and verify it is responsible for the trait

Types of Molecular Markers

- RFLP -restriction fragment length polymorphism
- SSR - small satellite repeats
- ISSR - Inter Simple Sequence Repeats

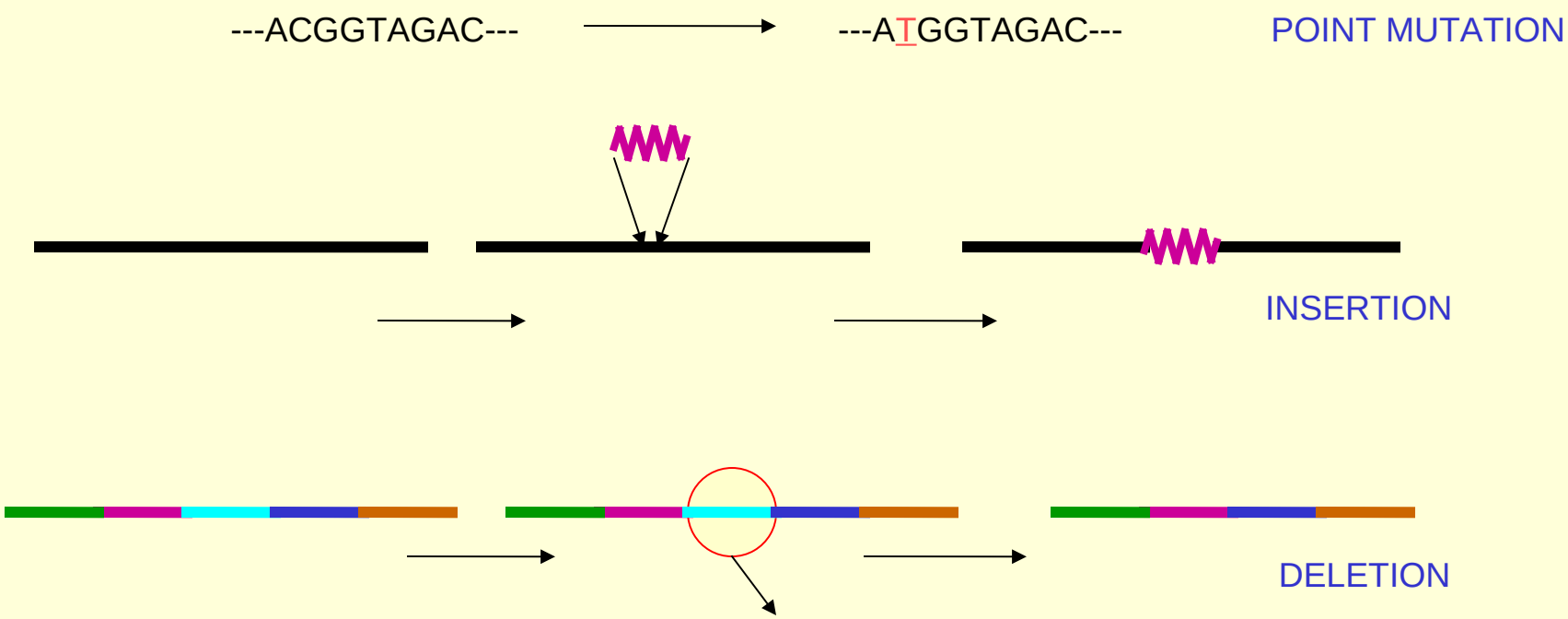
A molecular marker is any method that detects polymorphism (ie. DNA sequences differences) at a given locus between two parents in a genetic cross.

Sequence differences at the marker site are detected as differences in the length of fragments or the presence or absence of fragments created by PCR or restriction enzymes.

All marker methods do essentially the same thing: they allow us to detect the presence (1) or absence (0) of a short sequence, somewhere at a specific chromosomal location.

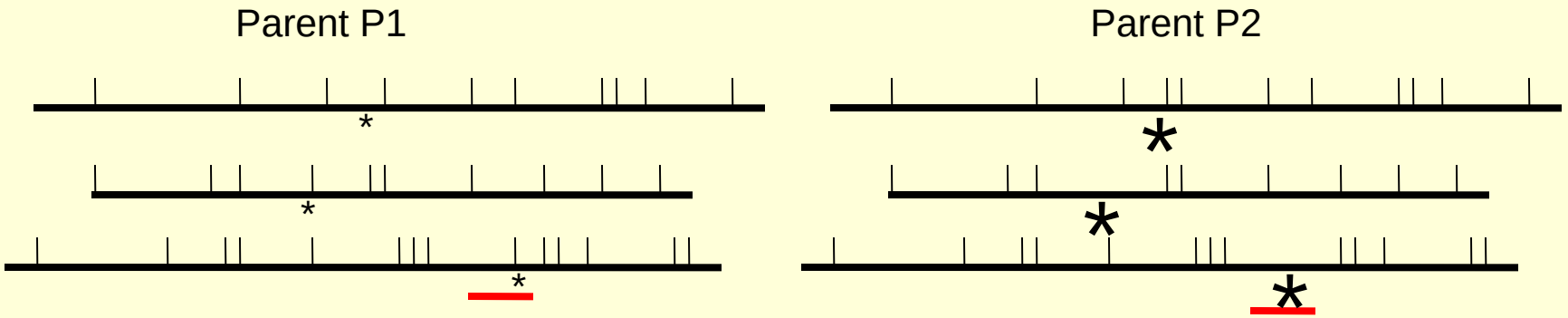
Polymorphisms

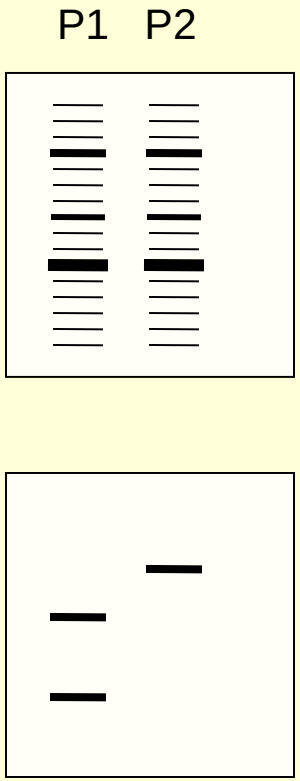
- Polymorphisms arise by mutations, insertions and deletions



Length polymorphisms are easy to understand for insertions and deletions but need to think of how length differences arise for point mutations

Each RFLP marker detects polymorphism at a single locus

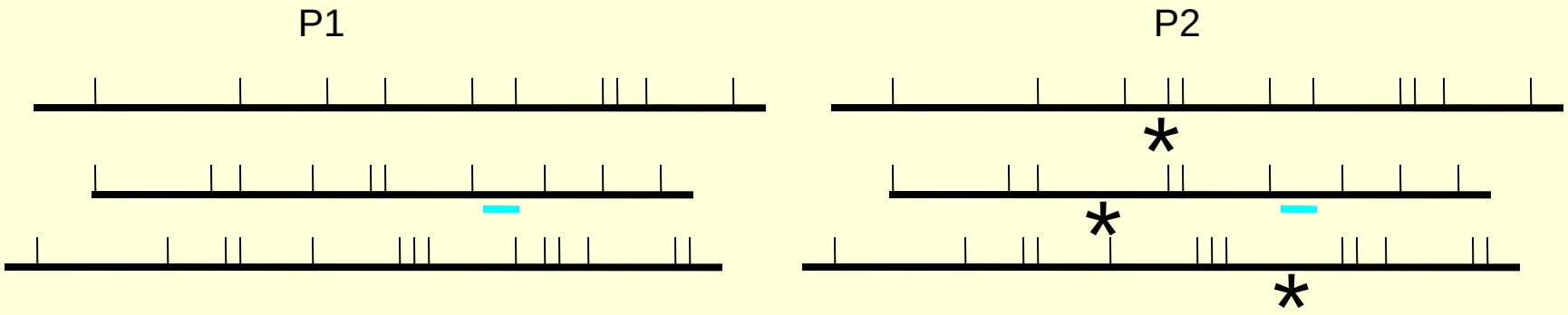




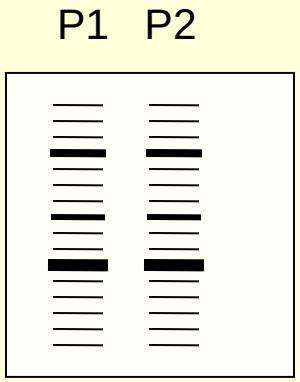
1. Total digestion of genomic DNA with restriction endonuclease EcoRI
2. Size fractionate DNA fragments on a gel
3. Blot DNA fragments to nylon membrane and denature
4. Hybridize with labeled DNA( ) probe representing a specific region of genome
5. Wash away unbound probe
6. Expose membrane with bound probe to film

Polymorphism exists between two parents for this restriction endonuclease and this probe

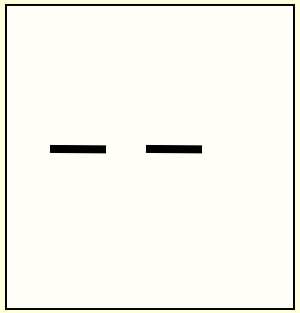
Monomorphic loci are uninformative



1. Digest total genomic DNA with restriction endonuclease EcoRI
2. Size fractionate DNA fragments on a gel



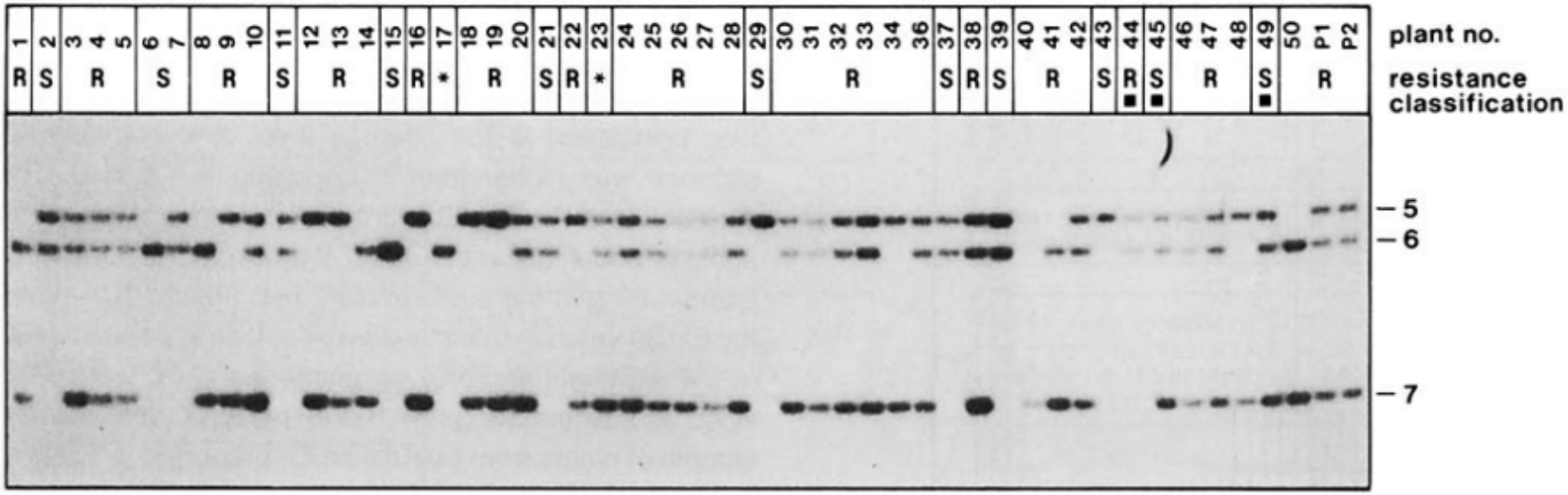
3. Blot DNA fragments to nylon membrane
4. Hybridize with labeled DNA( ) probe representing a specific region of genome
5. Wash away unbound probe
6. Expose membrane with bound probe to film



<u>NO polymorphism</u> is apparent between the two parents for this restriction endonuclease and this probe

Co-segregation of a band with a trait

Barzen et al. (1992) Plant J. 2:601-611.



Fragments 5 and 6 segregate independently	**1:1:1:1 ratio of R+:R-:S+:S-**
Fragment 7 co-segregates with resistance. Plants 17 and 23 were not screened for resistance. Plants 44, 45, and 49 are presumed to be recombinant between the marker hybridization site and the R locus.	R/+ : 34 parental R/- : 1 recombinant S/+: 3 recombinant S/- : 10 parental

Genomic DNA from 49 plants (1-34, 35-50) segregating for resistance to rhizomania in sugar beet was digested with TaqI and probed with marker GS3. The probe hybridized to 7 fragments - only the lower 3 are shown.

Co-segregation of a band with a trait

- Deviation of the phenotypic ratio from the expected 1:1:1:1 proves that the presence of the band is linked to Resistance.
- To calculate linkage distance, we need a larger number of individuals, and more information about the cross.

RFLP - Restriction Fragment Length Polymorphisms

- Different restriction enzymes will produce different patterns and have the potential to identify point mutations in their restriction sites
- Probes can come from any piece of a genomic or cDNA library as these will represent different parts of the genome

  eg a random selection of 100 pieces of DNA as probes will test 100 sites in the genome

Advantage:

- Produce co-dominant markers (see both alleles)
- Hybridization probe used in finding markers can be directly used in chromosome walking step

Disadvantage:

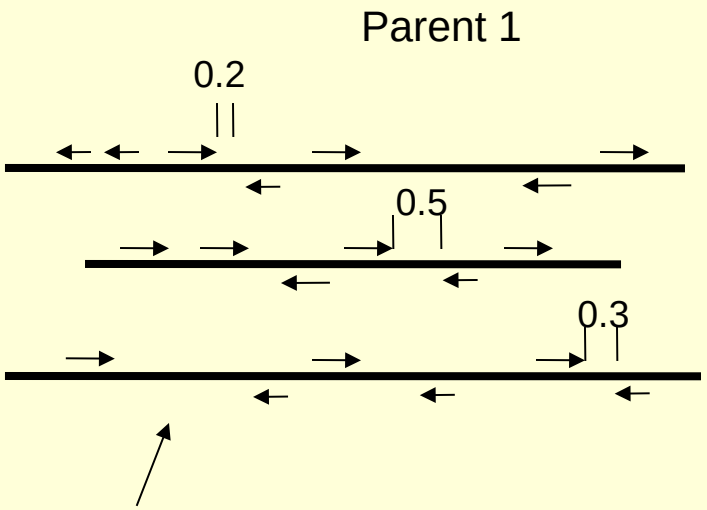
- Method is very labor intensive (Southern blots)
- Normally only get 1 or 2 potential markers per probe
- Must label probes to visualize differences

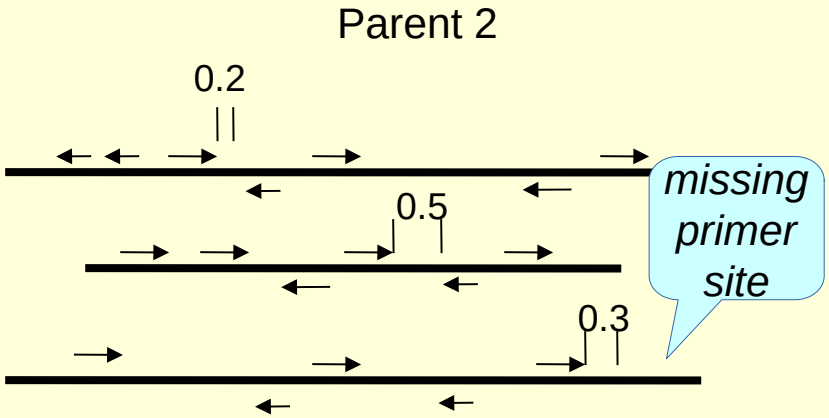
For these reasons, RFLPs are generally considered to be obsolete.

PCR-based markers

- Many variants of this approach eg. SRAP, RAPD, AFLP etc.
- Involves using PCR and short (10- 21 nt) primers to detect polymorphisms
- By random chance, any primer sequence should be found many times in a typical eukaryotic genome. Some of these priming sites will be found in close proximity, in opposite orientations.
- A point mutation will create or destroy a priming site in one of the parent genotypes - hence one of the priming sites necessary to create an amplified fragment will be lost.
- Small insertions or deletions between priming sites will appear as co-dominant markers. That is, because the two alleles give different sized bands, we can see both alleles in the heterozygote.

PCR markers

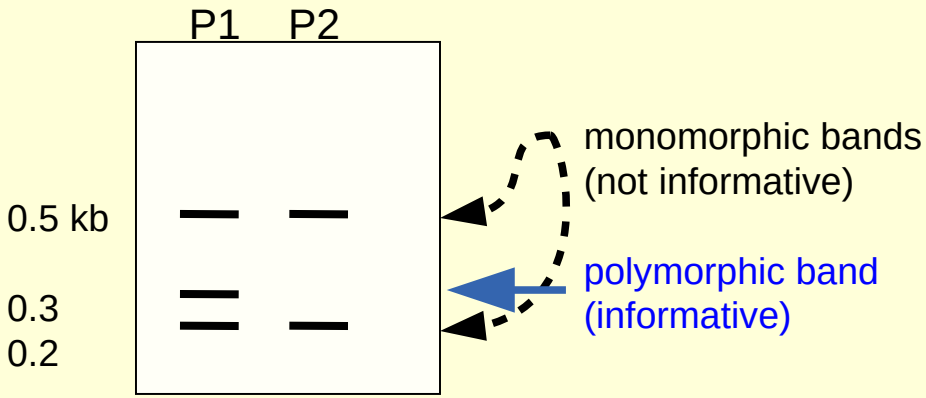




1. Isolate genomic DNA from each parent
2. Perform PCR
3. Size fractionate the PCR products on a gel

Individual small primers will find many sites to prime

Separation of amplified DNA on an agarose gel



The 0.3 kb band is polymorphic in the two genotypes and hence a potential marker

Potential markers must be linked to trait being sought

Example: screening for a marker that co-segregates with disease resistance in a plant population. Each row is a different locus.
Each column is a different plant
1 - presence of a band 0 - absence of a band

Parents		Progeny of cross and backcross of F1																				
P1	P2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	plant #
R	S	R	S	S	R	R	R	S	S	R	S	S	R	S	R	S	R	R	S	R	R	S	S	resistance
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RAPD45
0	1	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	1	0	1	RAPD62
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	RAPD49
1	0	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	1	0	0	1	1	RAPD87
1	0	1	0	0	1	1	1	0	0	1	0	0	1	0	1	0	1	0	0	1	1	0	0	RAPD145

Potential markers must be linked to trait being sought

P1	P2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	plant #
R	S	**R**	S	S	**R**	**R**	**R**	S	S	**R**	S	S	**R**	S	**R**	S	**R**	**R**	S	**R**	**R**	S	S	resistance
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RAPD45
0	1	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	1	0	1	RAPD62
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	RAPD49
1	0	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	1	0	0	1	1	RAPD87
1	0	**1**	0	0	**1**	**1**	**1**	0	0	**1**	0	0	**1**	0	**1**	0	**1**	0	0	**1**	**1**	0	0	RAPD145

RAPD62 - 16/22 distantly linked to resistance locus

RAPD87 - 8/22 not linked to resistance locus

RAPD145 21/22 closely linked to resistance locus

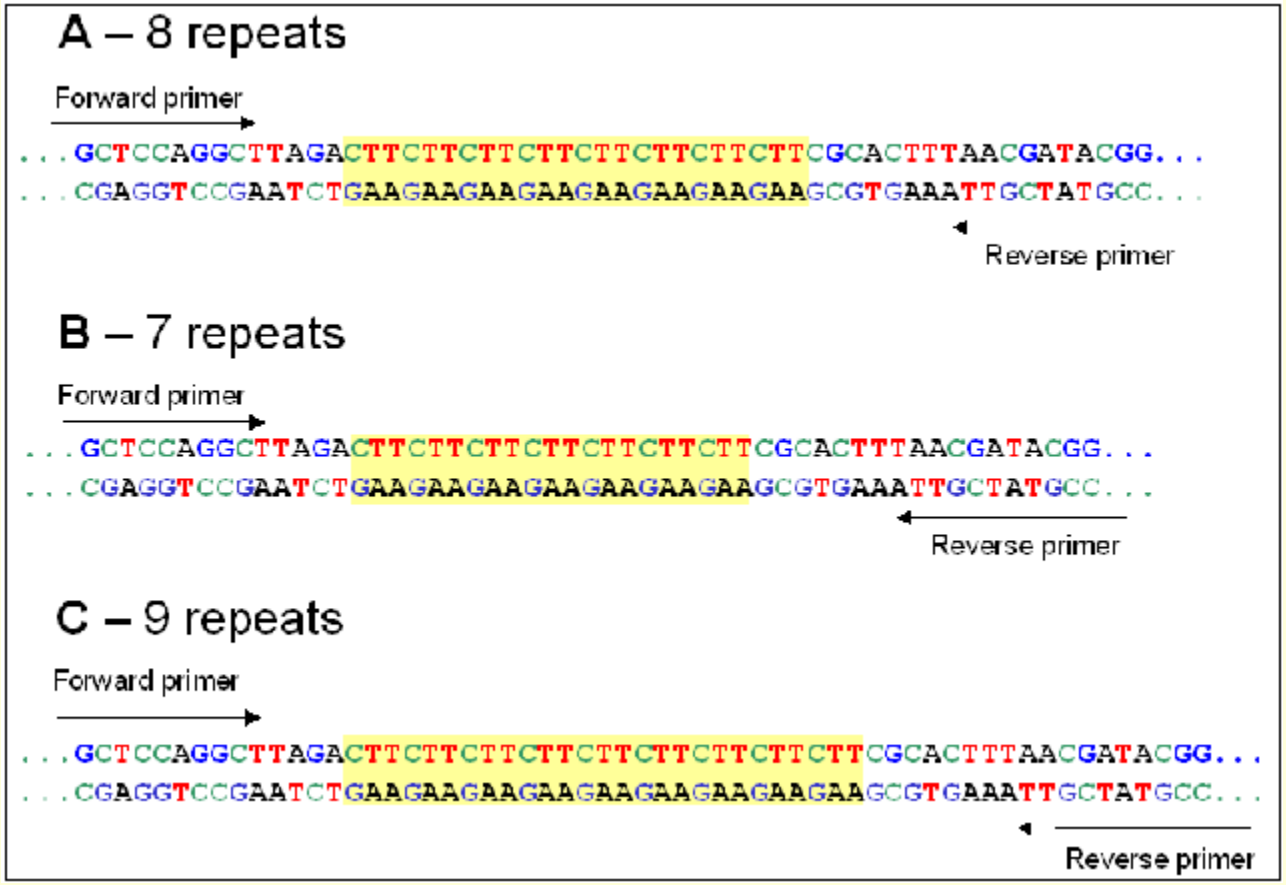
A high level of co-segregation of the trait polymorphism (resistance/susceptibility) with polymorphism associated with molecular marker RAPD145 indicates close linkage

The marker would need to be isolated from the gel, amplified, cloned and sequenced to be used for chromosome walking

In practice, need a few hundred segregating progeny to measure linkage distance.

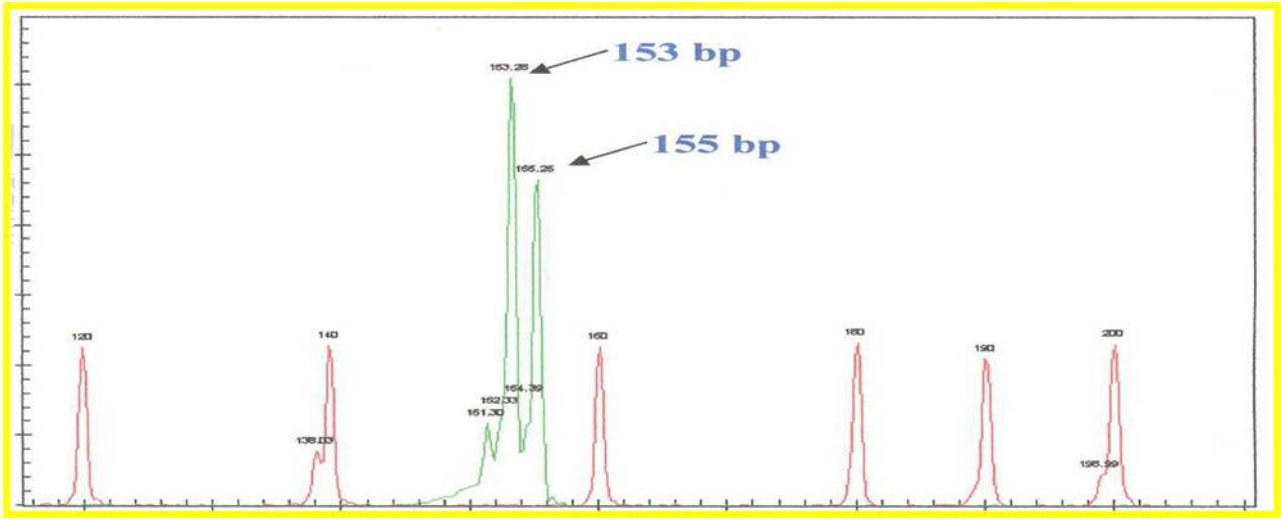
SSR - small satellite repeats (microsatellites)

- short sequences consisting of very short repeat units in tandem arrays.
- interspersed with unique DNA throughout eukaryotic chromosomes.
- At any locus, there are usually several alleles, each having different numbers of repeat units
- Each allele generates a distinct band in PCR.



SSR - small satellite repeats (microsatellites)

To visualize the different bands produced in PCR, PCR products are run on gels. Fluorescently-tagged PCR products SSRs are typically run on capillary gels, and the products detected as peaks of fluorescence as bands migrate past a fluorescence detector.



from: http://hilo.hawaii.edu/depts/epscor/Workshops.php

In this example, two alleles for a single microsatellite locus are seen, one at 153 bp and the other at 155 bp. Because both bands are seen, we know that the individual from which the products were amplified was heterozygous for these two alleles, at that locus.

Non-anchored ISSRs (Inter Simple Sequence Repeats)

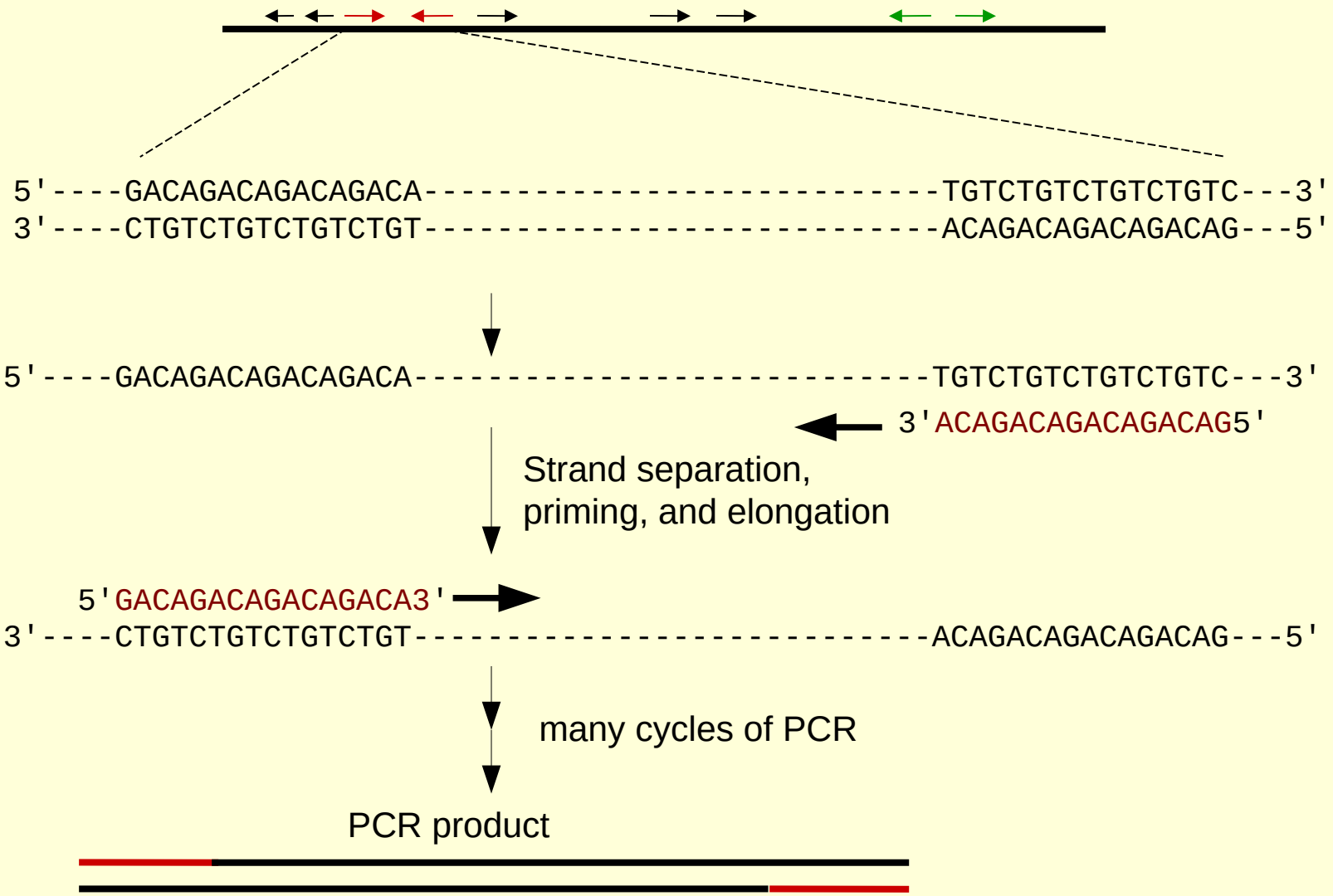
Basis: Many genomes contain short satellite repeats interspersed among other repetitive or single-copy sequences across the genome. At a small percentage of sites, two copies of these repeats occur in inverted orientation, separated by a few hundred bp.

Microsatellites with many different repeat units are found in eukaryotic genomes.
Simply create primers made from short repeats, and test each with your genome of interest.

Primers: multimers of short repeat units
ex. $(CAG)_5$ = CAGCAGCAGCAGCAG
$(CAA)_5$= CAACAACAACAACAA
$(GACA)_4$ = GACAGACAGACAGACA

Ref: Branchard M, Bornet B (2001) Plant Mol. Biol. Reporter 19:209-215.

Example of primer (GACA)$_4$



Non-anchored ISSRs (Inter Simple Sequence Repeats)

Like microsatellite markers:

- each band corresponds to an individual locus containing a microsat.
- polymorphism can be based on loss of a band, or changes in repeat number

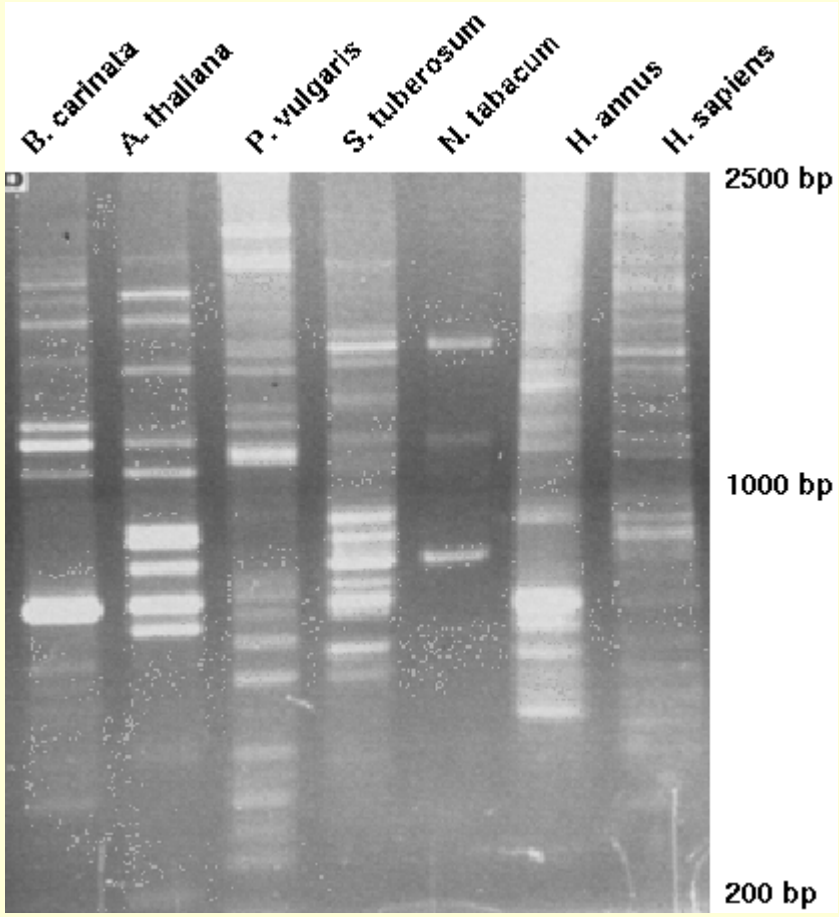
Unlike microsatellite markers:

- primer is microsat. seq. itself, rather than flanking unique sequence
- uses a single primer; assumes two copies of microsat. in inverted orientation
- no prior knowledge of genome is required.

Non-anchored ISSRs (Inter Simple Sequence Repeats)

- A single primer may give different bands in different species.
- It is <u>not</u> safe to assume that two bands of the same size, in different species, represent the same genetic locus in both species. In other words, the bands are probably not homologous.

example: $(CAA)_5$ primer



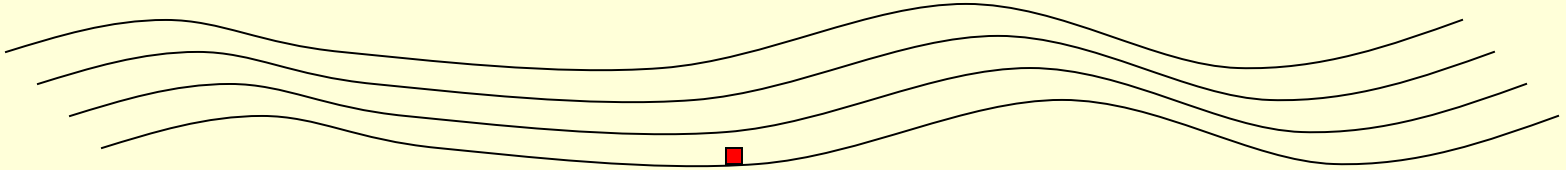
Ref: Branchard M, Bornet B (2001) Plant Mol. Biol. Reporter 19:209-215.

From close marker to the trait gene

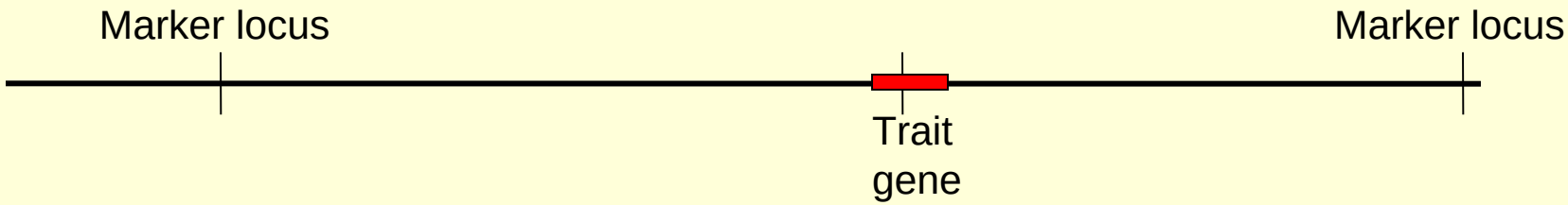
- By any of these methods you will end up with a piece of DNA representing the molecular marker locus near the trait gene

- This can be used as a probe to screen a genomic library to find a large piece of DNA corresponding to the marker region

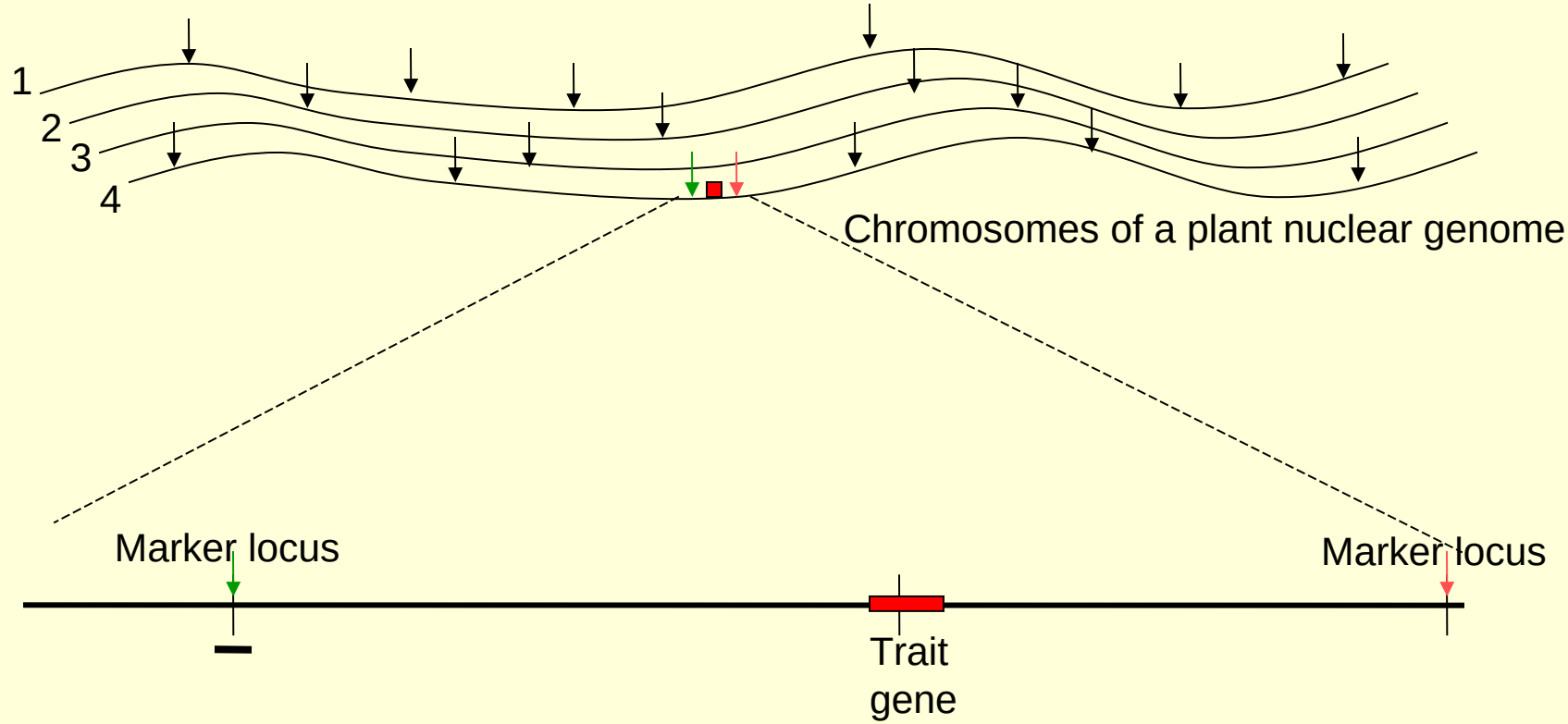
Chromosome Walking from marker to trait gene



Chromosomes of a plant nuclear genome

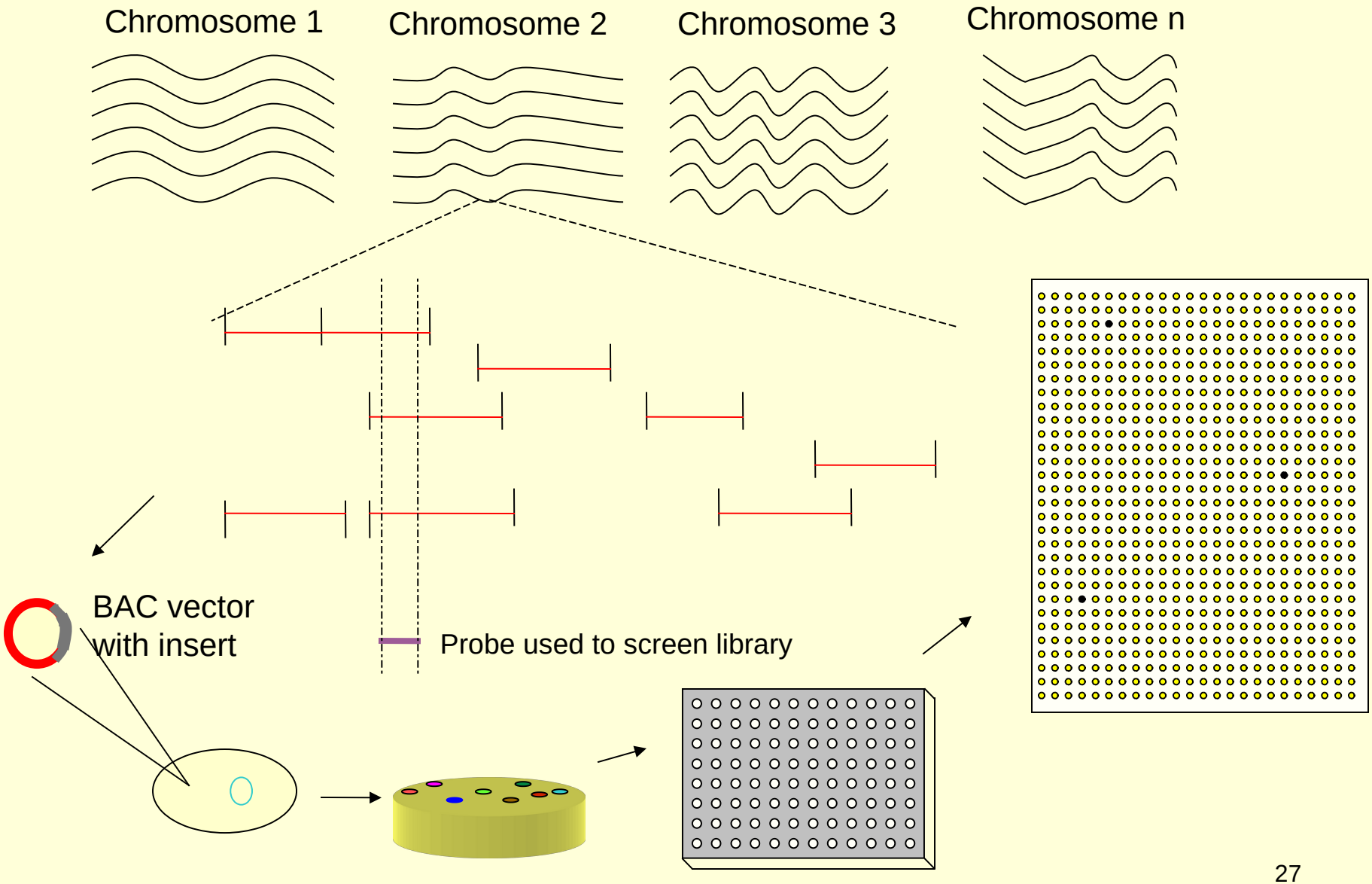


Chromosome Walking from marker to trait gene

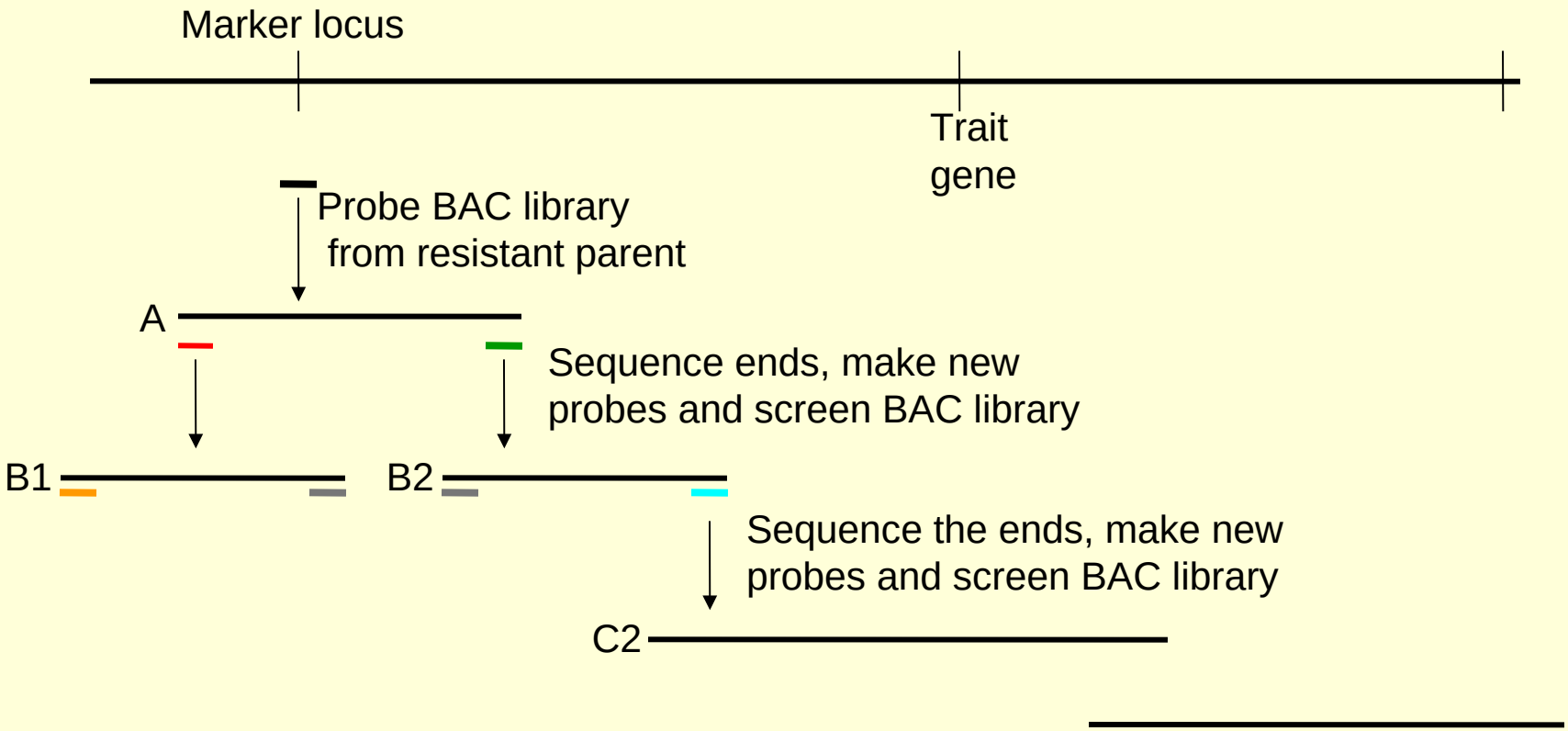


Chromosome walking involves finding overlapping pieces (each piece is a step) of the genome from a library of large fragments (BAC library) derived from resistant parent.

Genome exists in BAC library as series of overlapping fragments



Chromosome Walking from marker to trait gene



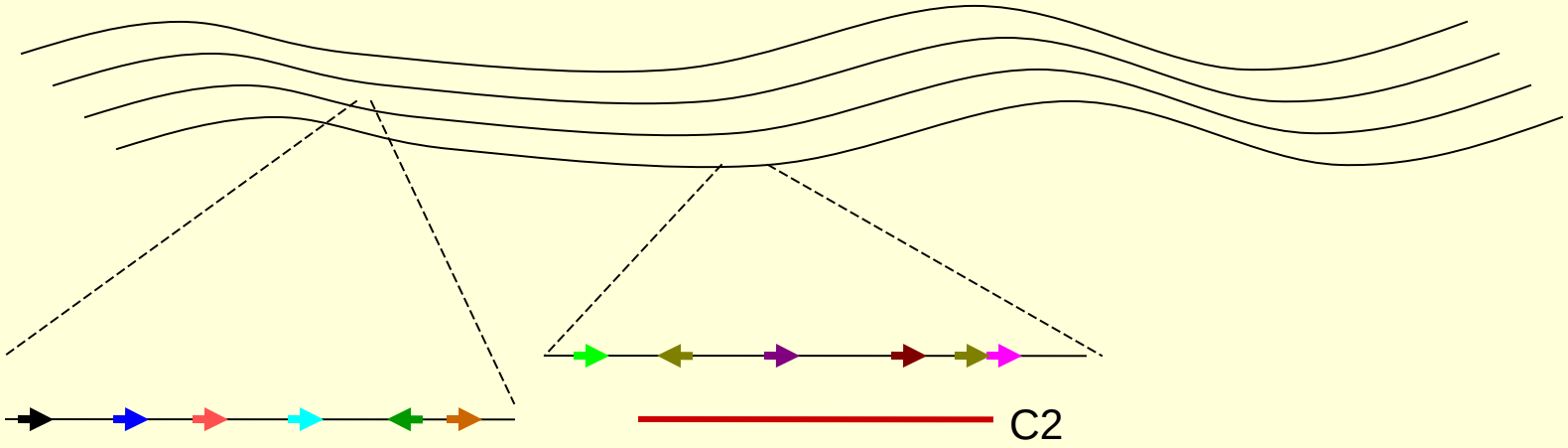
Can find one or two genomic pieces that likely contain the trait gene eg. C2

Locating the gene in the selected genomic fragments

- Prepare a cDNA library from the resistant parent during the expression of resistance
  - Expectation the trait gene is being expressed.

- Use C2 fragment as a probe to screen the cDNA library.
  - Find all the genes which occur on the C2 fragment that are being expressed.
  - These are the potential trait genes and all would be sequenced and potential functions identified from sequence databases.

- Narrow the number of possible genes to a few

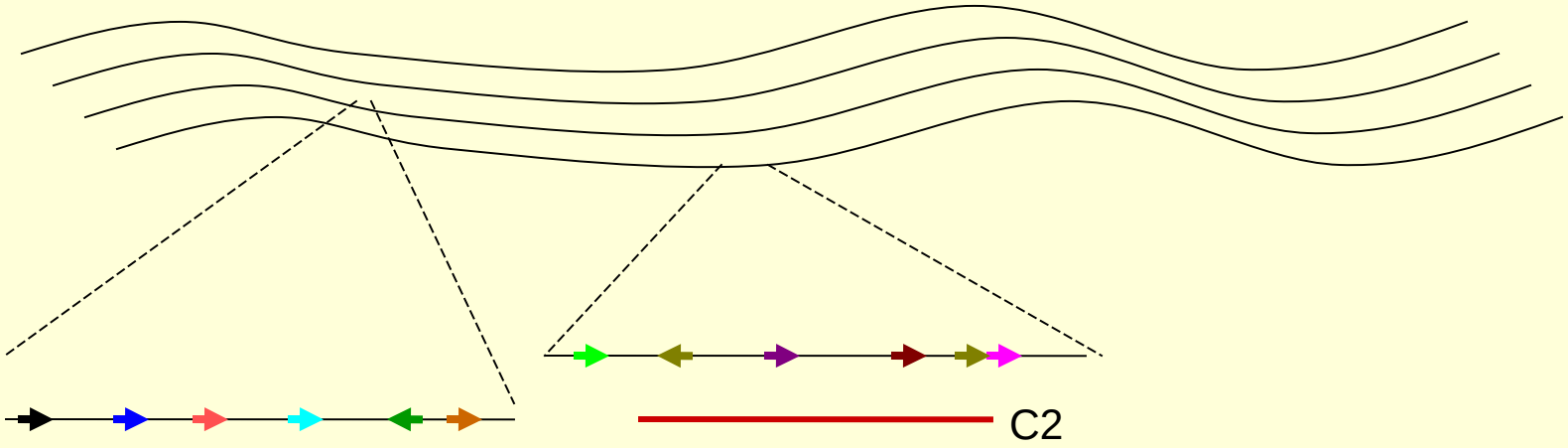


BAC clone believed to have trait gene

Isolate mRNA from tissues of plant expressing resistance

1000's of mRNAs

cDNA library



BAC clone believed to have trait gene

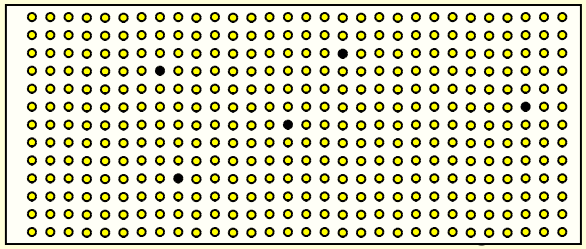
Isolate mRNA from tissues of plant expressing resistance

mRNAs

cDNA library

Label C2 probe and screen library

cDNA library on a membrane



Locating the gene

- Prepare a cDNA library from the resistant parent during the expression of resistance
  - Expectation the trait gene is being expressed.

- Use C2 fragment as a probe to screen the cDNA library.
  - Find all the genes which occur on the C2 fragment that are being expressed.
  - These are the potential trait genes and all would be sequenced and potential functions identified from sequence databases.

Narrow the number of possible genes to a few based on what we know about our gene and its expression.

Verification – proving a particular gene produces the trait

- Insert each potential trait gene under the control of a constitutive promoter (always on) into the susceptible parent (lacks the trait) cells and regenerate whole plants.

- If the resultant plants of one of the inserted genes exhibits the trait then that is your trait gene. Proceed to full characterization

- Use cDNA of selected gene to go back to final genomic clone and get the endogenous promoter as well - sequence.

- If you want to know difference between resistant and susceptible plant isolate allele from susceptible plant and compare to resistant allele

Review - Finding a gene

- From cDNA or genomic library when information about the gene or gene product is known

- By positional cloning when only the trait is known

  Based on discovery of a molecular marker (by RFLP,SRAP,ISSR or other methods) which are closely linked to the trait locus, followed by chromosome walking to the trait locus and identification of the gene.

The final step is to verify that the gene confers the function in transgenic plants. This would require cloning the functioning gene (eg. resistance gene) from one parent and transforming it into another plant line that is missing the function (eg. susceptible plant)

.