

# Marqueurs moléculaires et cartographie de QTL

# 1. Les marqueurs moléculaires

# Notion de marqueur génétique















- Marqueur génétique = caractère variable mesurable
- Permettant de caractériser/identifier un individu
- Leur transmission héréditaire peut être suivie au cours des générations

# Marqueurs génétiques: propriétés désirables

- Très polymorphes
- Reproductibles
- Codominants
- Distribués régulièrement le long du génome
- Discriminants
- Non soumis à l'influence de l'environnement
- Peu coûteux
- Faciles à mesurer

# La première génération de marqueurs génétiques

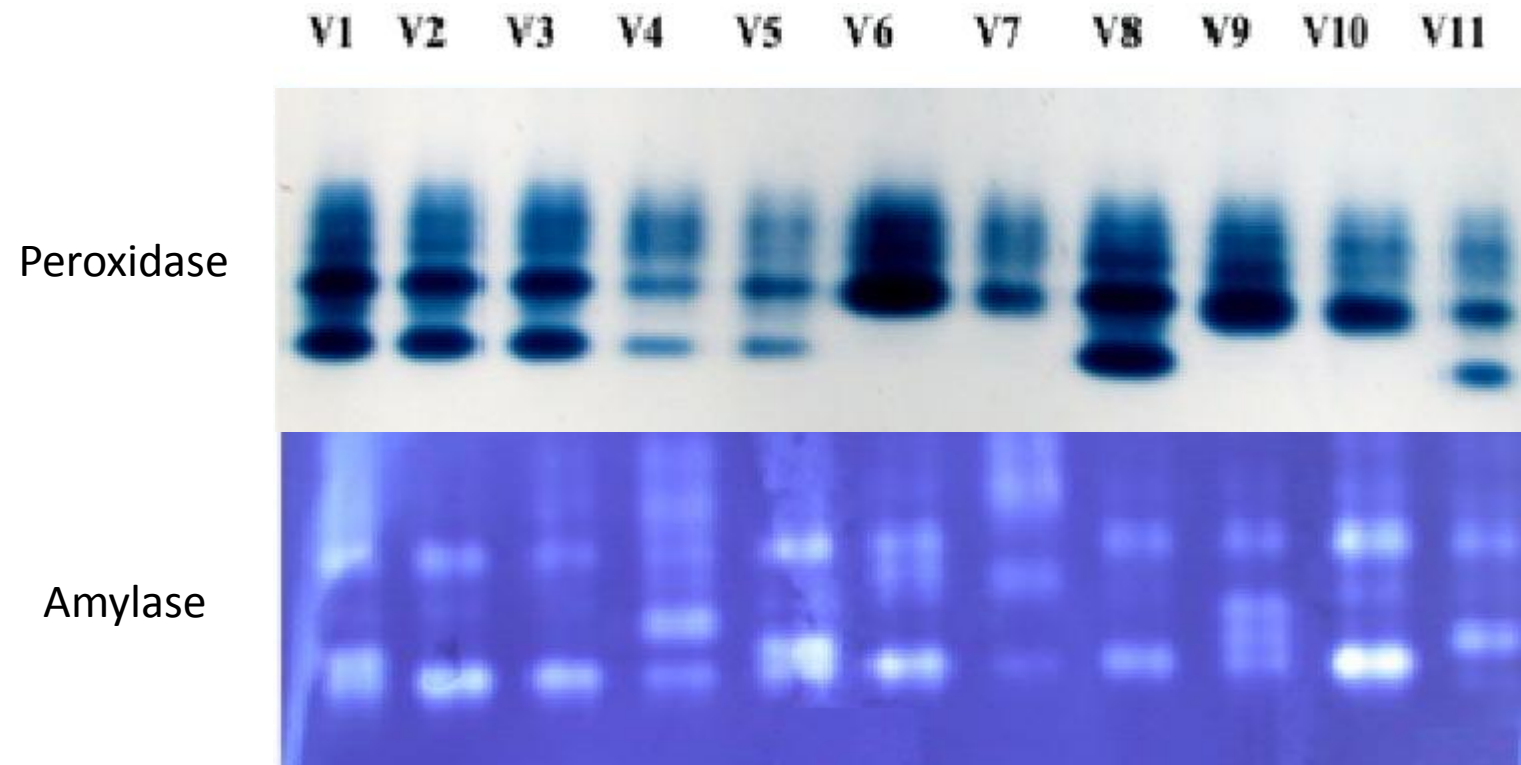
- **Les marqueur phénotypiques (morphologiques):** caractères observables, variables, monogéniques, peuvent être utilisés pour identifier un individu
- Inconvénients: nombre limité, influencés par l'environnement
- Il faudra développer autres types de marqueurs..... Les marqueurs biochimiques

Character	Dominant Trait	×	Recessive Trait	F <sub>2</sub> Generation Dominant:Recessive	Ratio
Flower color	Purple 	×	White 	705:224	3.15:1
Flower position	Axial 	×	Terminal 	651:207	3.14:1
Seed color	Yellow 	×	Green 	6022:2001	3.01:1
Seed shape	Round 	×	Wrinkled 	5474:1850	2.96:1
Pod shape	Inflated 	×	Constricted 	882:299	2.95:1
Pod color	Green 	×	Yellow 	428:152	2.82:1
Stem length	Tall 	×	Dwarf 	787:277	2.84:1

# Les marqueurs biochimiques

- Protéines: Enzymes, protéines de réserve
- Electrophorèse + coloration histochimique spécifique de la protéine en question
- Isozymes: différentes formes d'une enzyme canalisant la même réaction mais codées par différents *loci*
- Allozymes: différentes formes d'une enzyme codées par différents allèles d'un locus
- Ici, isozymes au sens large
- Protéines souvent utilisées:

# Les isozymes (isoenzymes)



**Isozyme profiles of oats cultivars**

# Les isozymes (isoenzymes)

- Inconvénients:
  - Soumis à l'influence de l'environnement
  - Nombre limité: relativement peu d'essais biochimiques disponibles pour révéler des enzymes

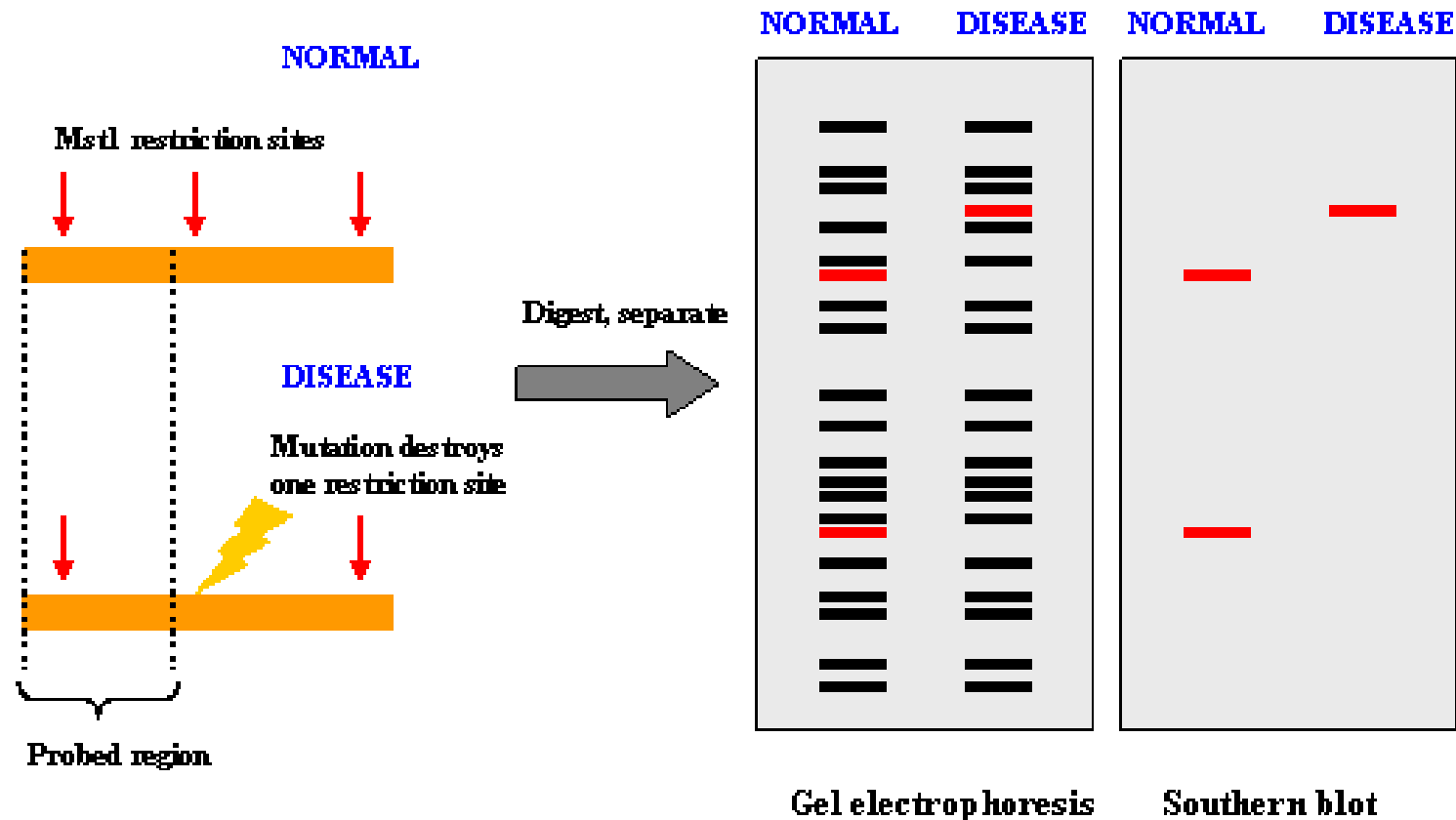


# Marqueurs moléculaires

- Basés sur la molécule d'ADN
- Avantages:
  - Non influencés par l'environnement
  - Potentiellement, en nombre illimité
  - Permet d'accéder l'origine de toute variation mesurable

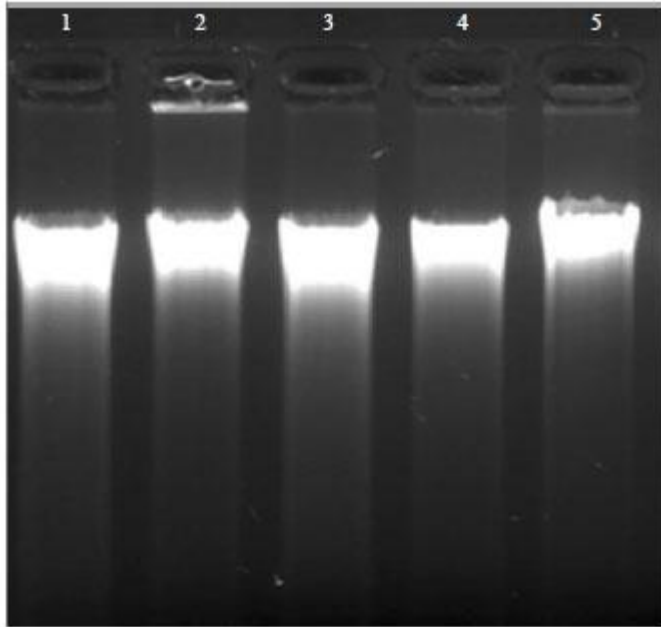
# Polymorphisme de longueur de fragment de restriction

## *Restriction Fragment Length Polymorphism (RFLP)*

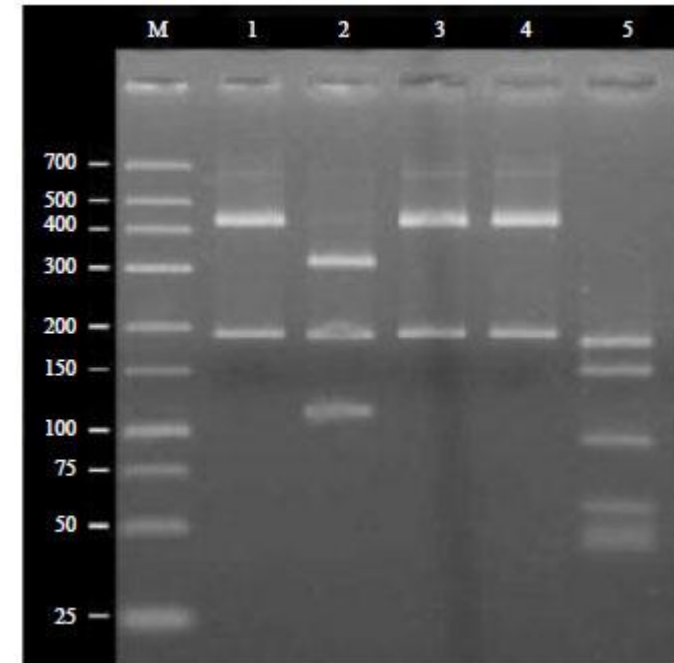


# Polymorphisme de longueur de fragment de restriction

## *Restriction Fragment Length Polymorphism (RFLP)*



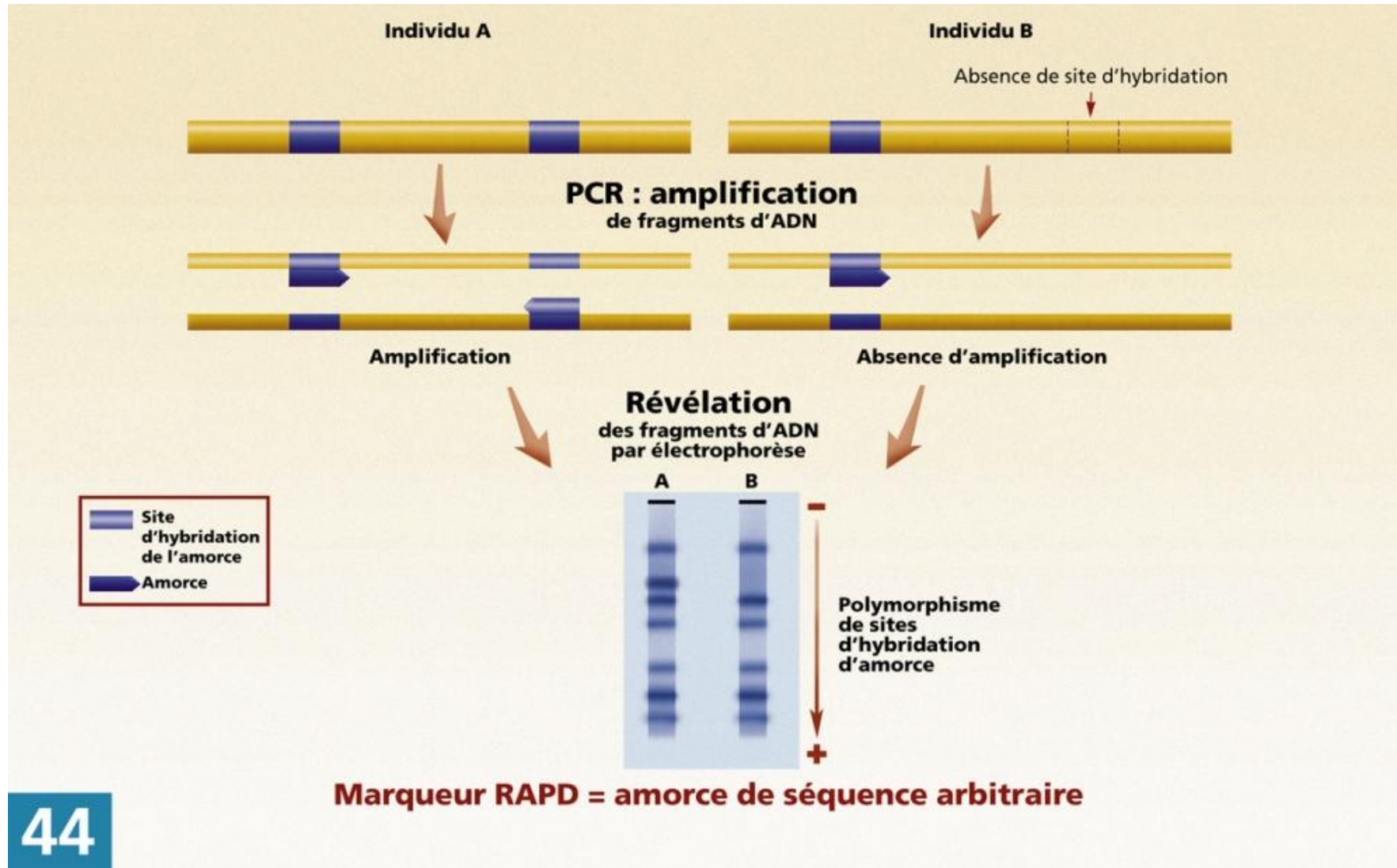
Digested genomic DNA



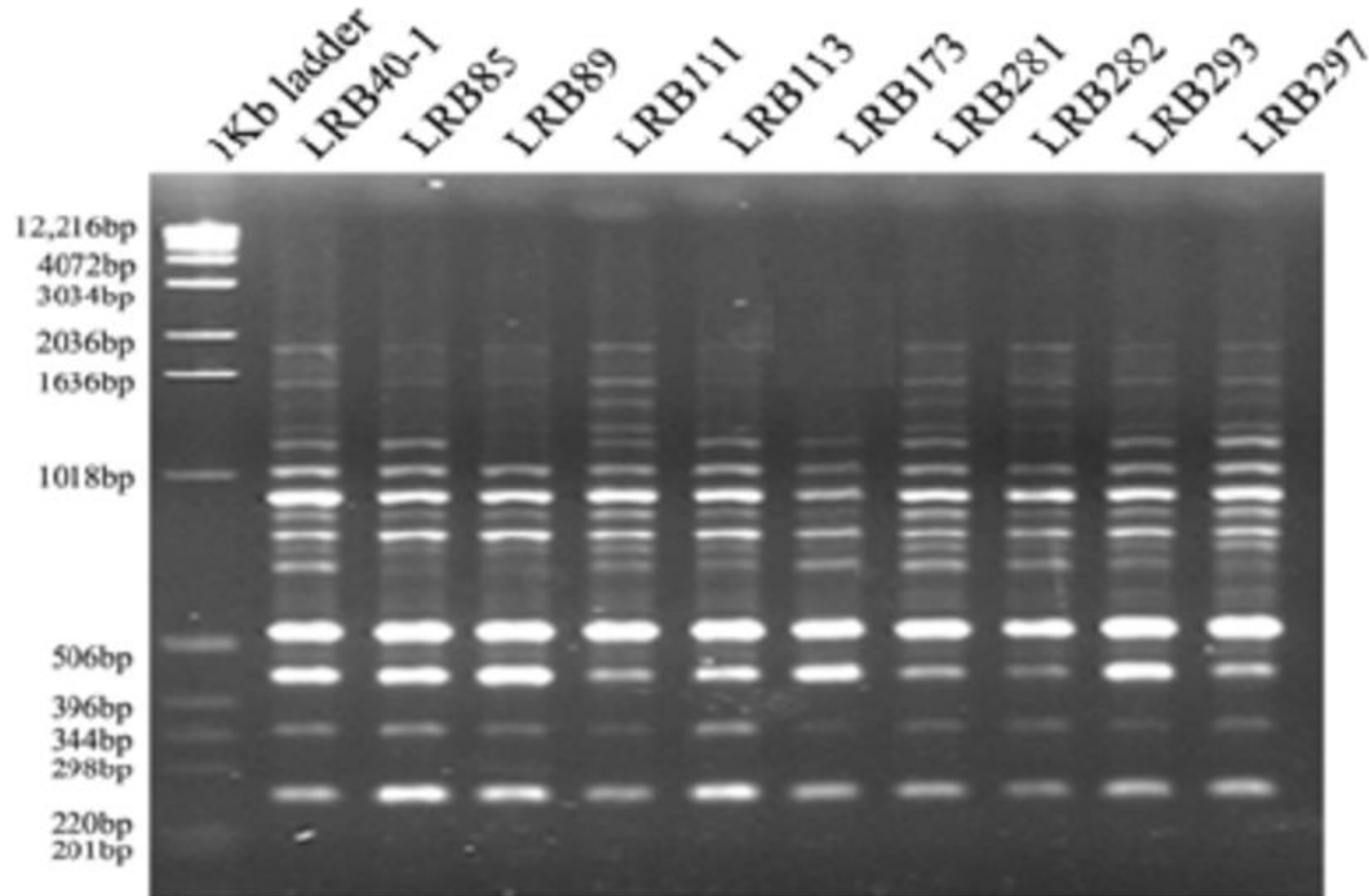
Revelation

# ADN polymorphe amplifié au hasard

## *Random Amplified Polymorphic DNA (RAPD)*



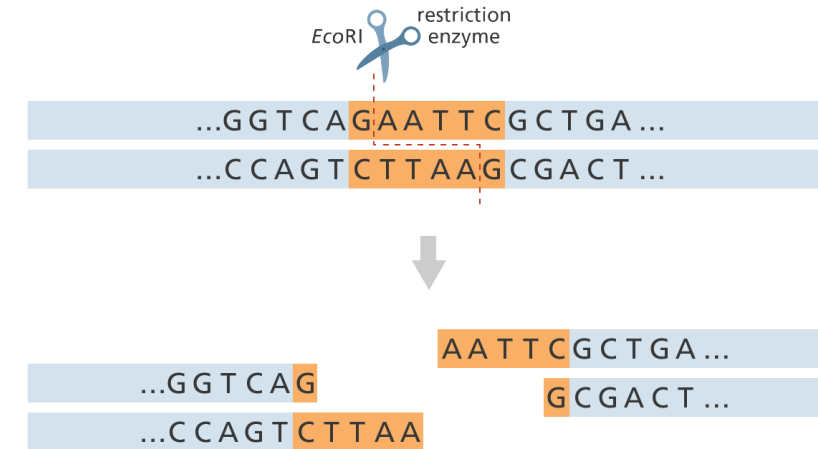
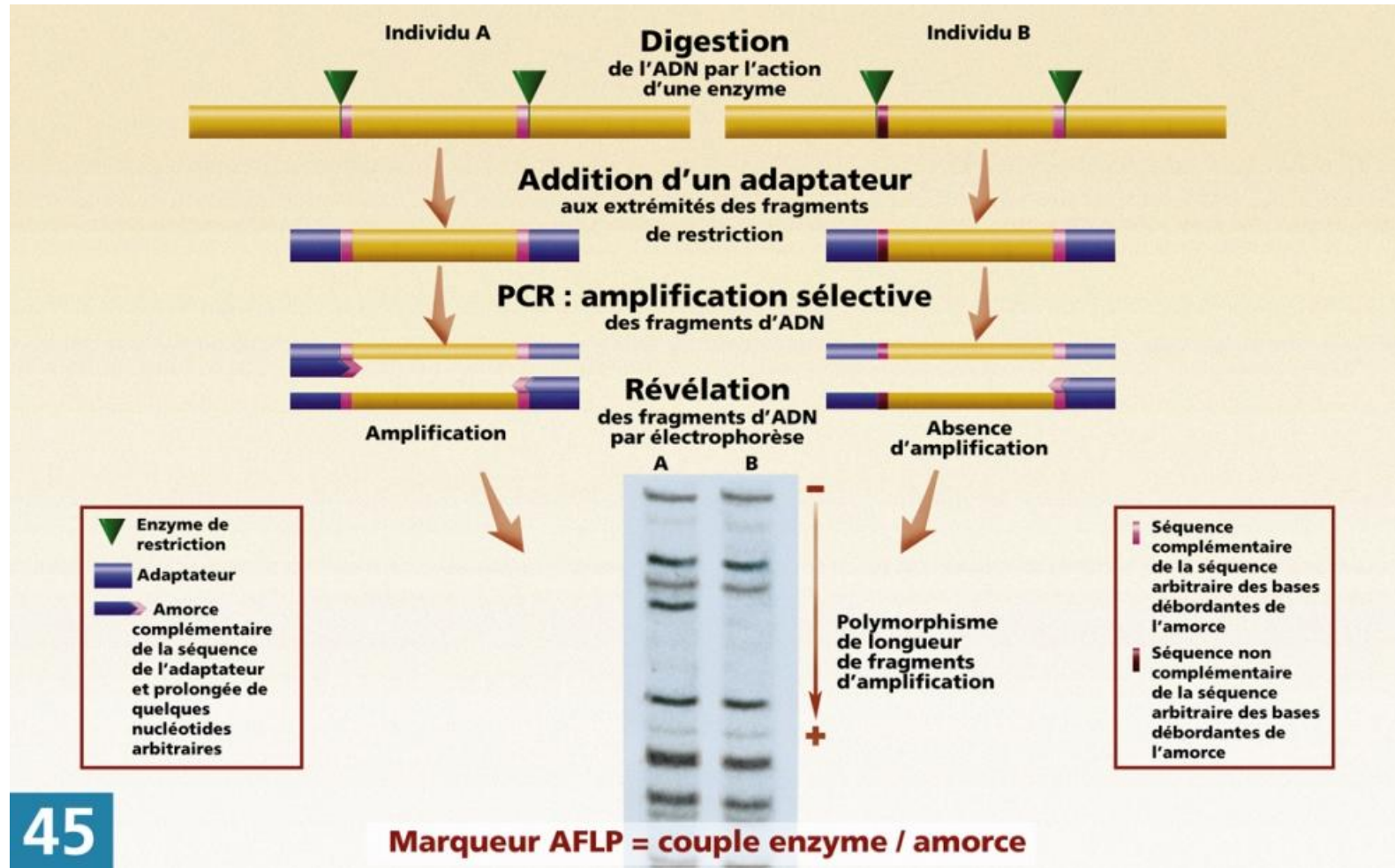
# ADN polymorphe amplifié au hasard *Random Amplified Polymorphic DNA (RAPD)*



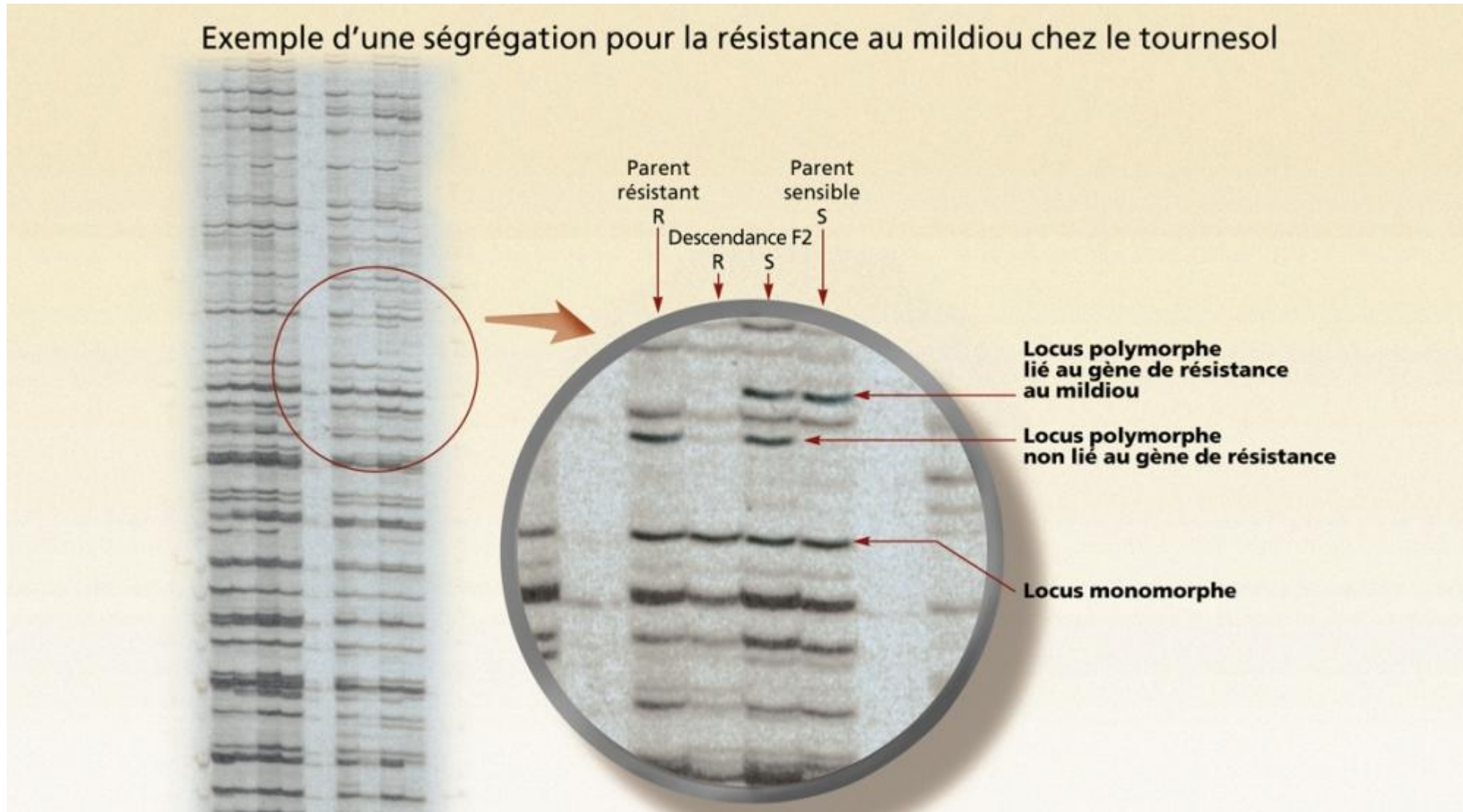
RAPD marker profiles of 10 landraces of *V. umbellata* generated by primer OPBB1

# Polymorphisme de longueur des fragments d'amplification

## *Amplification Fragment Length Polymorphism (AFLP)*



# Polymorphisme de longueur des fragments d'amplification *Amplification Fragment Length Polymorphism (AFLP)*



# Les marqueurs microstallites

## *Simple Sequence Repeats (SSR)*

ACTGTCG**ACACACACACACAC**GCTAGCT (AC)<sub>7</sub>  
TGACAGCT**TGTGTGTGTGTGT**GCGATCGA

ACTGTCG**ACACACACACACACAC**GCTAGCT (AC)<sub>8</sub>  
TGACAGCT**TGTGTGTGTGTGTGT**GCGATCGA

ACTGTCG**ACACACACACACACACACAC**GCTAGCT (AC)<sub>10</sub>  
TGACAGCT**TGTGTGTGTGTGTGTGTGT**GCGATCGA

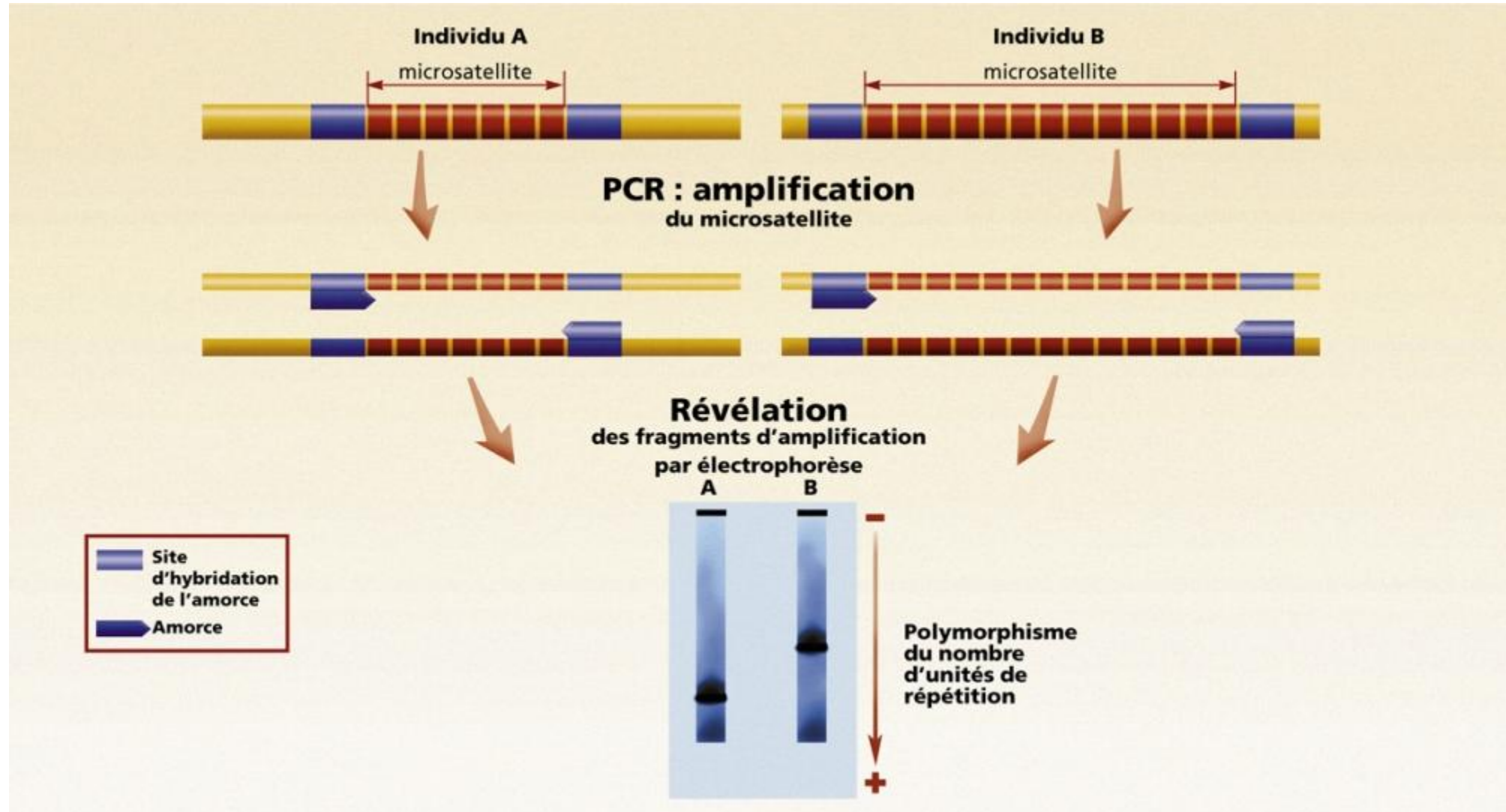
ACTGTCG**ACACACACACACACACACACACAC**GCTAGCT (AC)<sub>12</sub>  
TGACAGCT**TGTGTGTGTGTGTGTGTGTGTGT**GCGATCGA

Polymorphisme de nombre d'unités de répétition



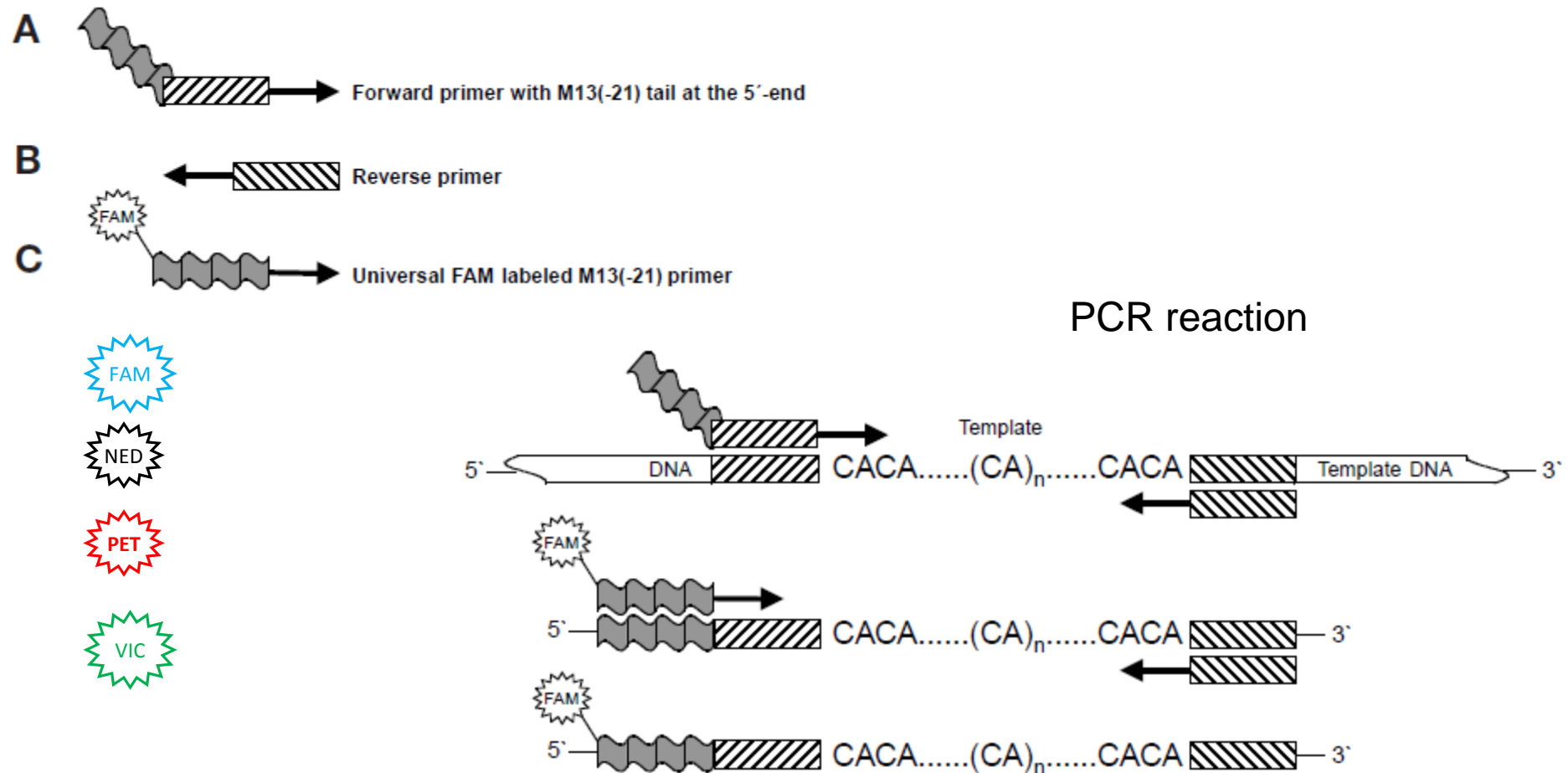
# Les marqueurs microstallites

## *Simple Sequence Repeats (SSR)*



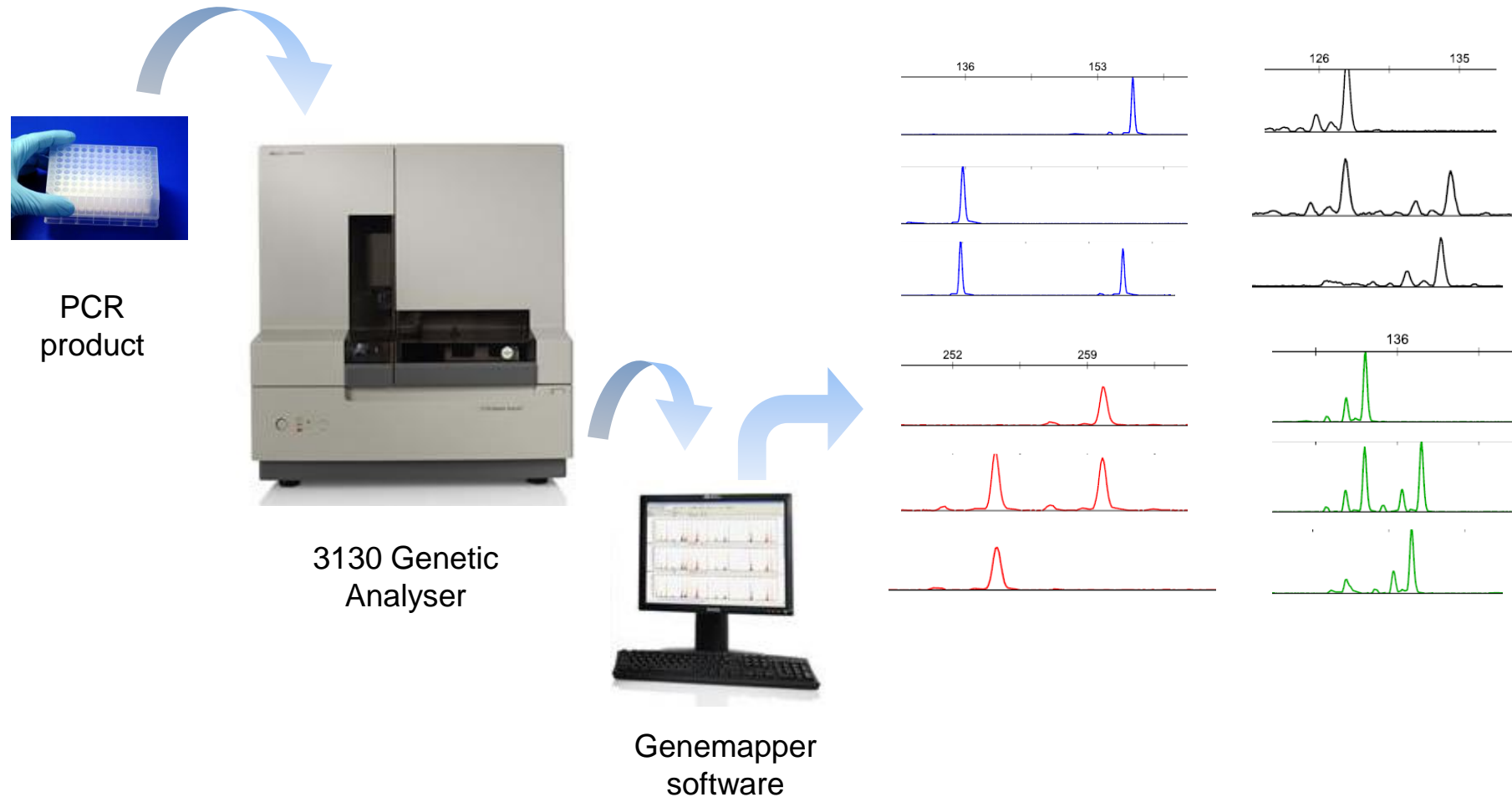
# Analyse de fragment en utilisant le séquenceur

## M13 -tailed PCR with fluorescent primer labelling



# Les marqueurs microstallites *Simple Sequence Repeats (SSR)*

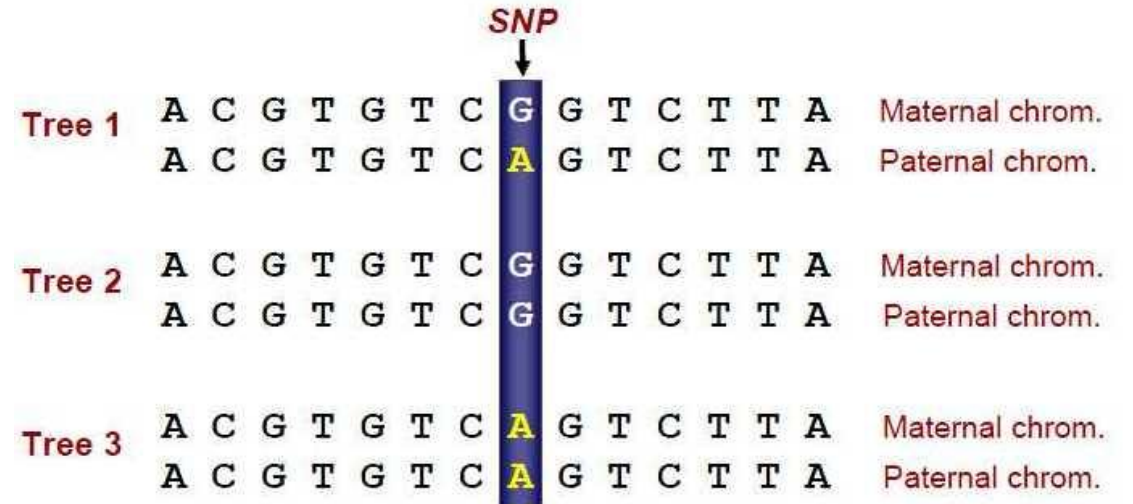
## PCR product visualization



# Single Nucleotide Polymorphism (SNP)

- **Avantages**

- Abondance
- Aptitude à l'automatisation



# Single Nucleotide Polymorphism (SNP)

- **Techniques de génotypage de première génération**

- Basées sur deux étapes:

- 1. Découverte:** re-séquençage ou criblage des bibliothèques de séquences préexistantes

- 2. Génotypage:** développement d'une plateforme de génotypage (microarray) pour un ensemble des SNP découverts (position, fréquences alléliques)

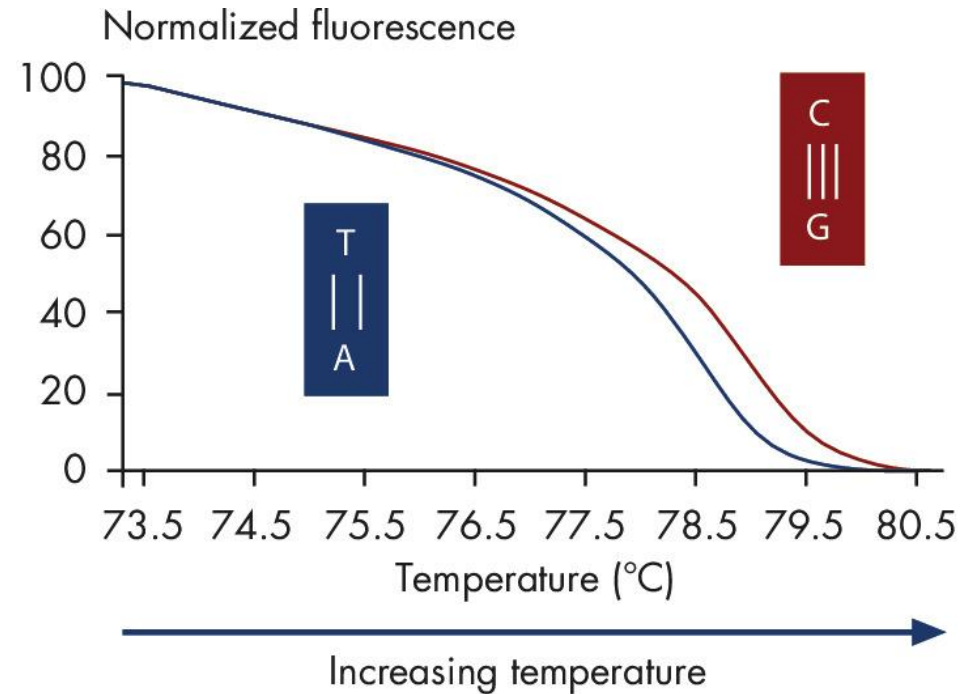
- Plusieurs technologies:

- Génotypage de SNP individuels: CAPs, HRM

- Génotypage multiplex à grande échelle: Les puces à ADN (DNA microarrays)

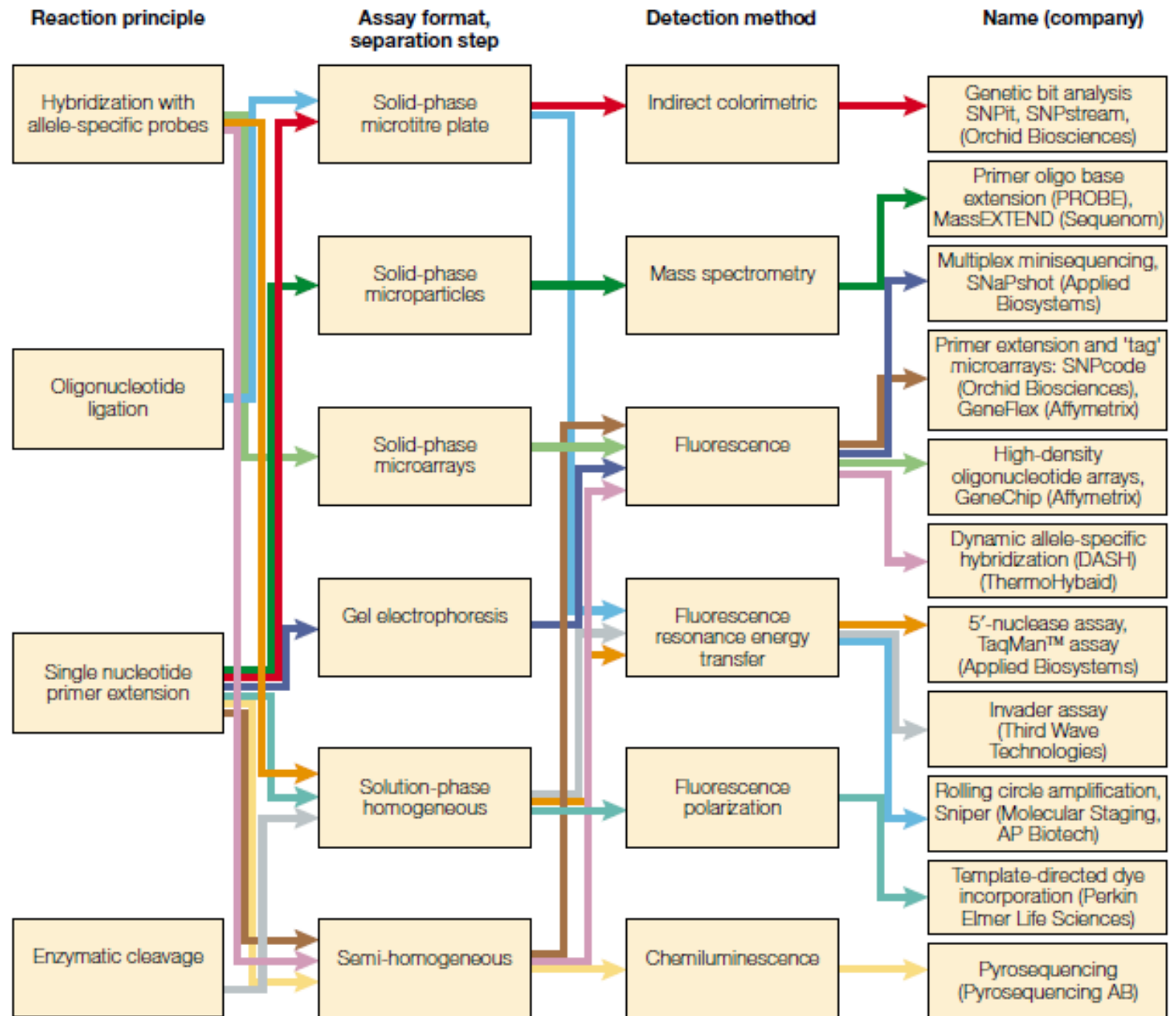
# High Resolution Melting (HRM)

- Basé sur la dissociation de l'ADN double brin sous l'effet de la T°c
- Dépend de la GC % et de la distribution des nucléotides
- Éléments nécessaires:
  - Produits d'amplification
  - Un fluorophore intercalaire spécifique à l'ADN double brin
  - Thermocycleur, détecteur de fluorescence et système de visualisation

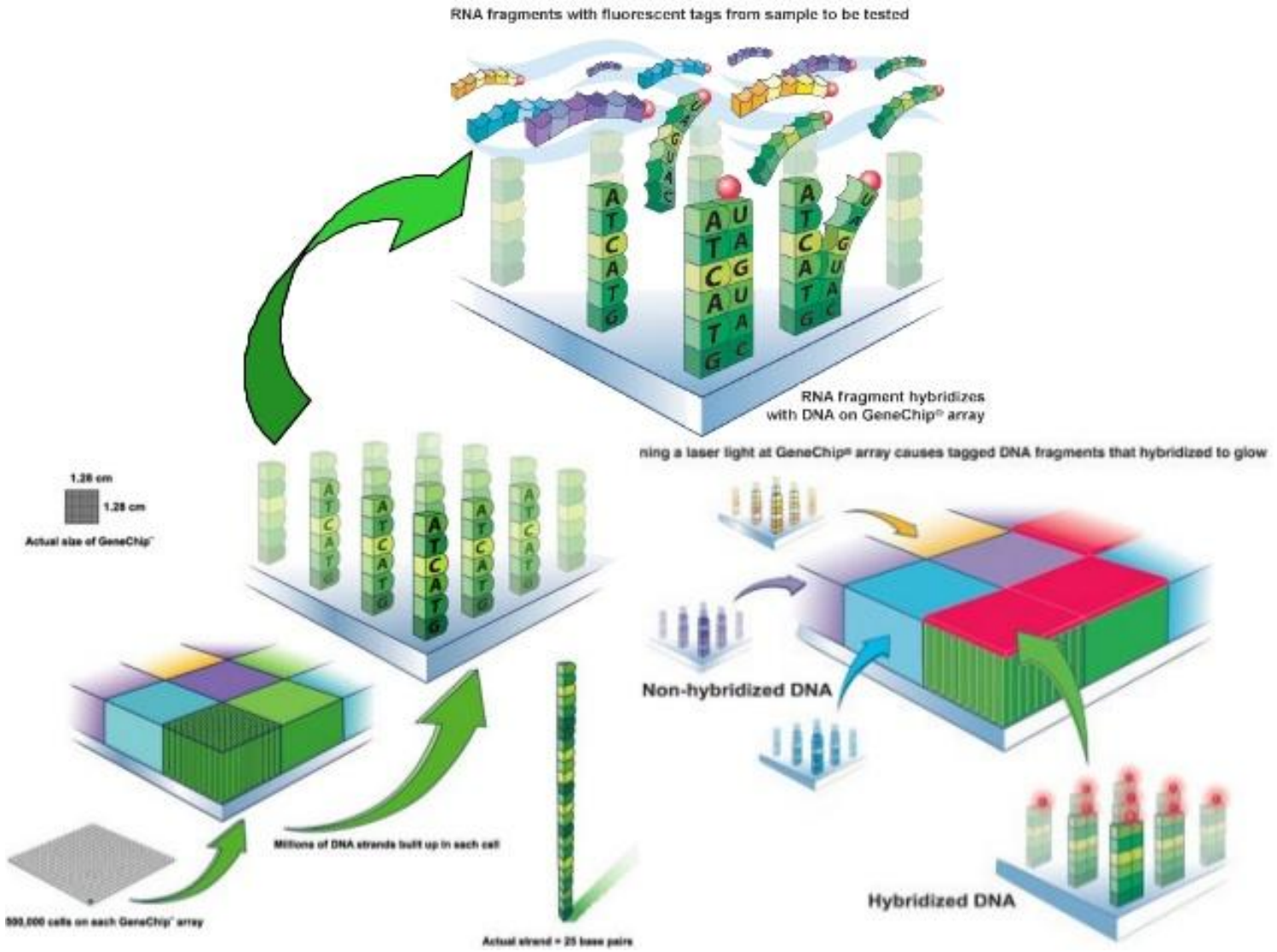


<https://www.qiagen.com/dz/resources/technologies/hrm/principle%20of%20hrm%20technology/>

# Génotypage à grande échelle (SNP microarrays)



# Génotypage à grande échelle (SNP microarrays)





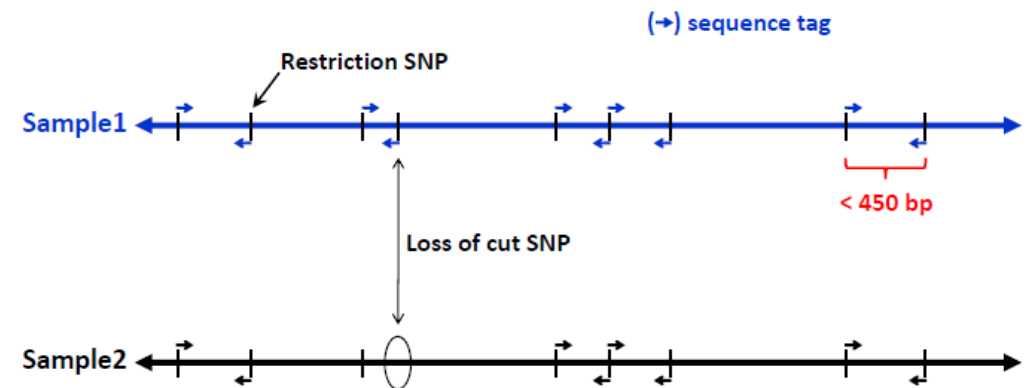
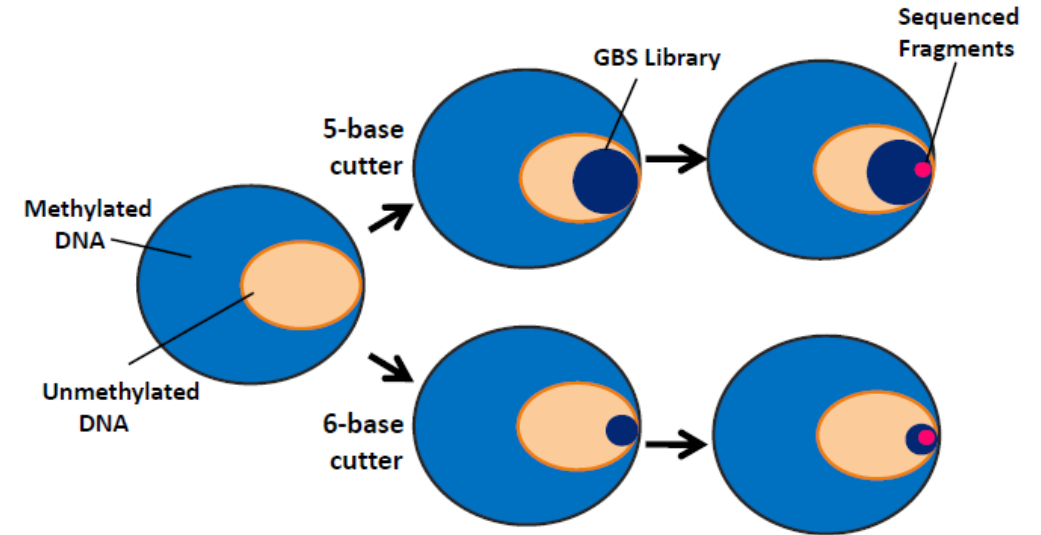
# Single Nucleotide Polymorphism (SNP)

- **Techniques de génotypage de deuxième génération**
  - Basées sur une seule étape: découverte et génotypage simultanés par séquençage
  - Next génération sequencing technologies (NGS)
  - Ex: **GBS** (Genotyping By Sequencing) basée sur la technologie de séquençage de Illumina

# Genotyping By Sequencing (GBS)

## ■ Principles:

- Reduction of genome complexity using methylation sensitive REs which filter out repetitive genomic fraction.
- Sequencing focused to the ends of restriction fragments (**64 bp**).



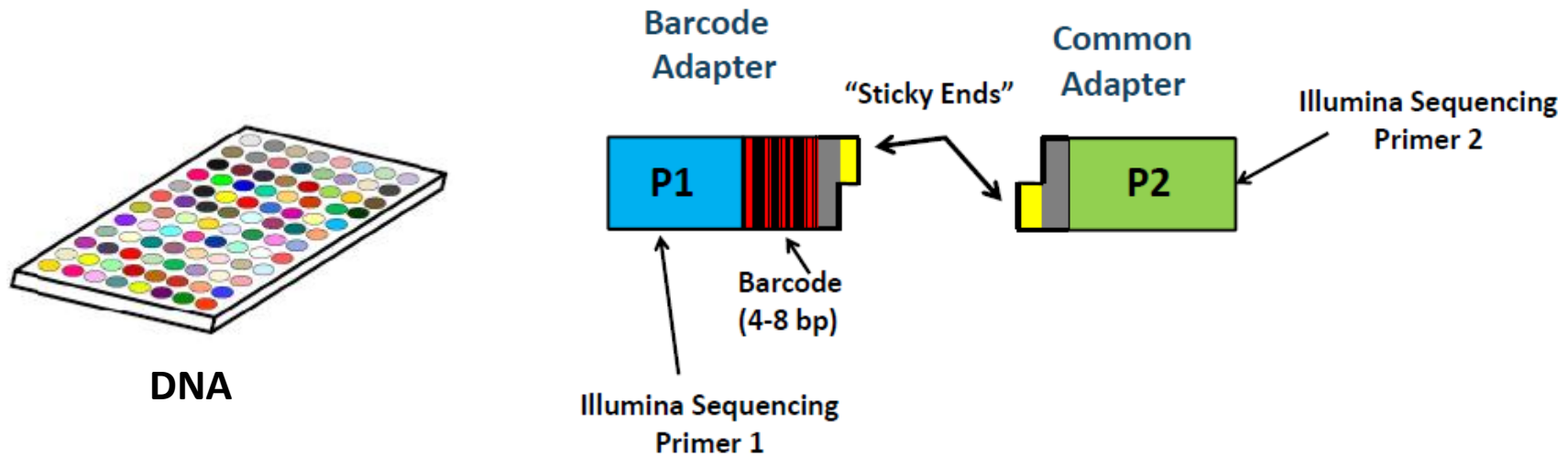
# Genotyping By Sequencing (GBS)

## 1. DNA sequencing

- Library preparation: DNA digestion + adaptors ligation
- Cluster generation: Bridge PCR
- Sequencing: Reversible terminator sequencing

## 2. Bioinformatics processing of sequence data for SNP identification

# Library preparation: needed materials



DNA

Barcode Adapter

Common Adapter

"Sticky Ends"

Illumina Sequencing Primer 2

Barcode (4-8 bp)

Illumina Sequencing Primer 1

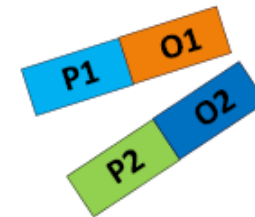
## Restriction Enzymes

*ApeKI* 5' G CWGC 3'

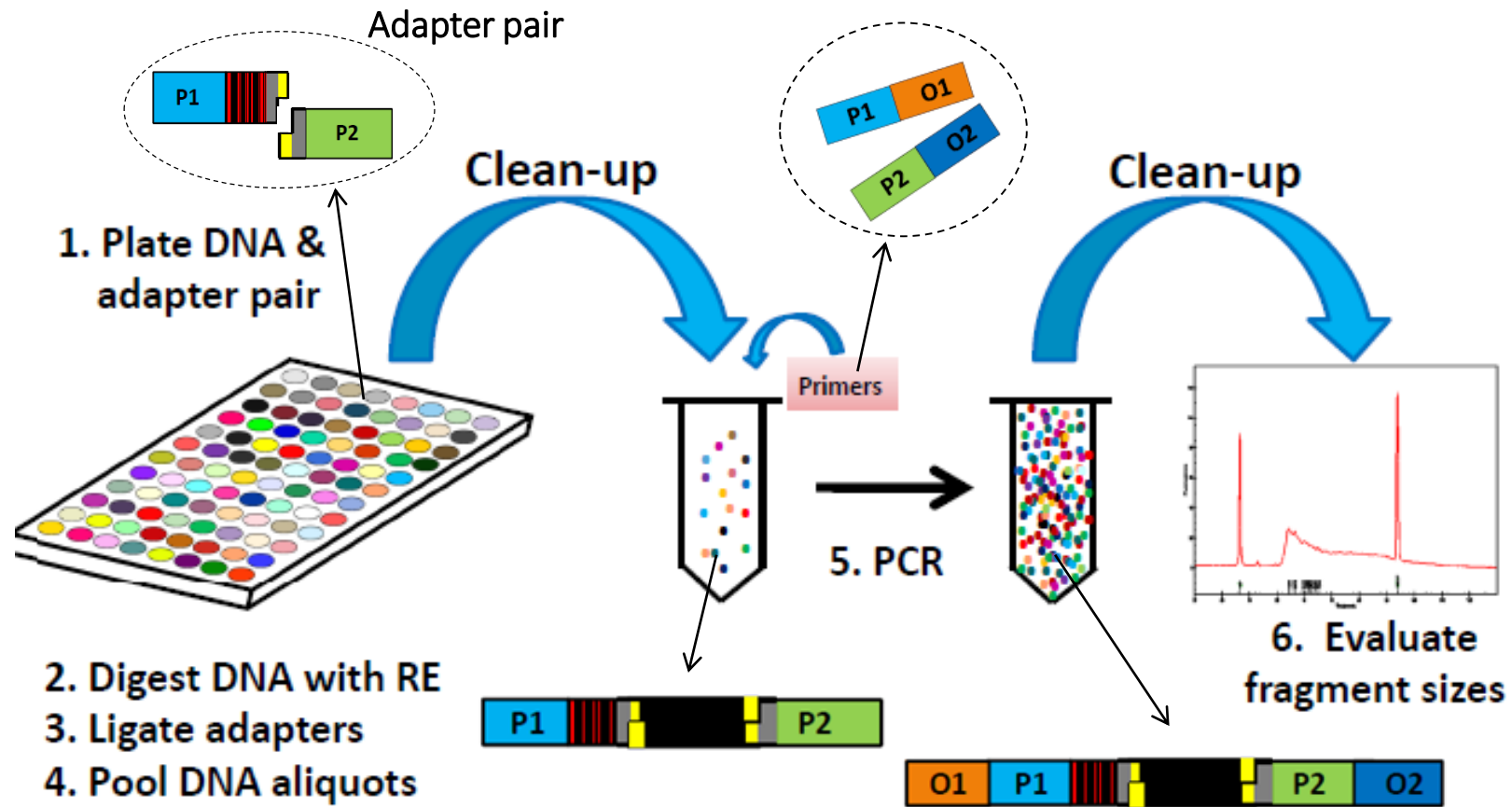
*PstI* CTGCA G

*EcoT22I* ATGCA T

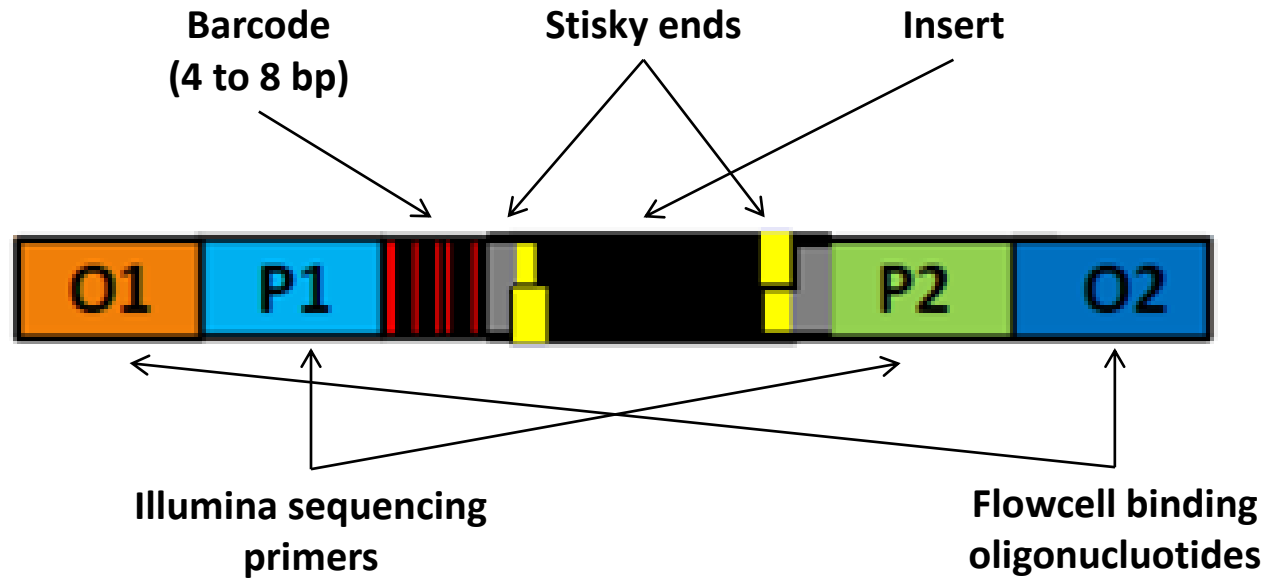
## PRC primers



# Library preparation workflow

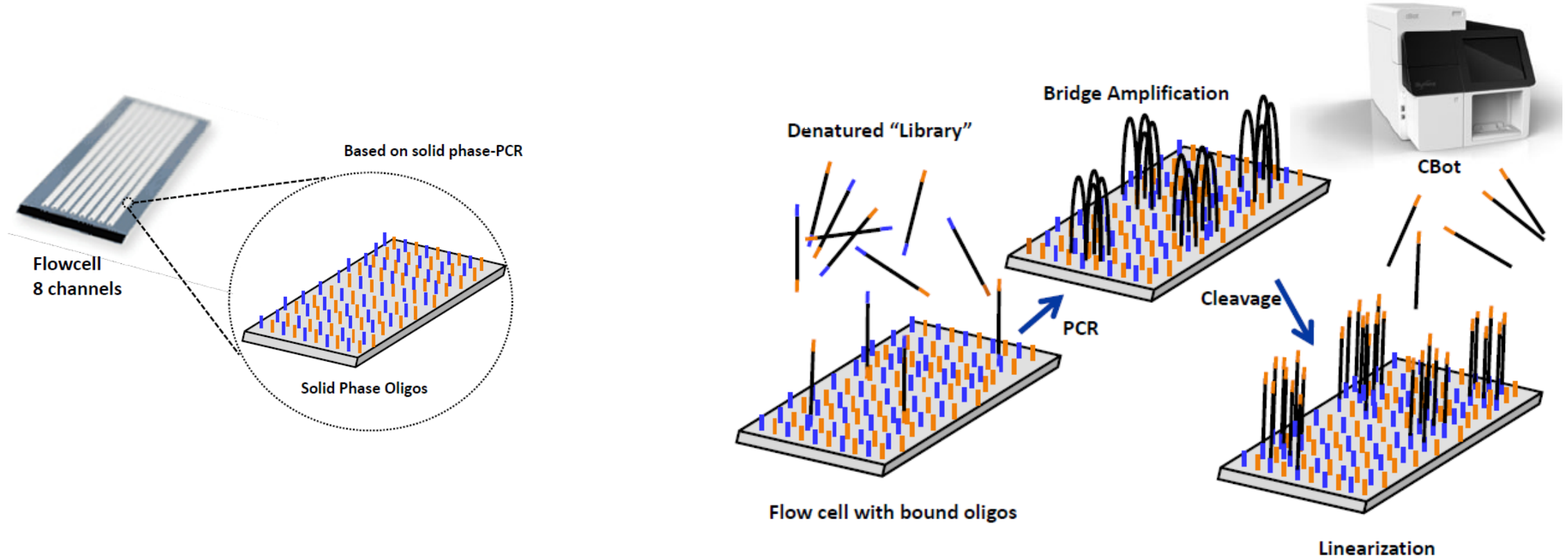


# Sequencing fragment

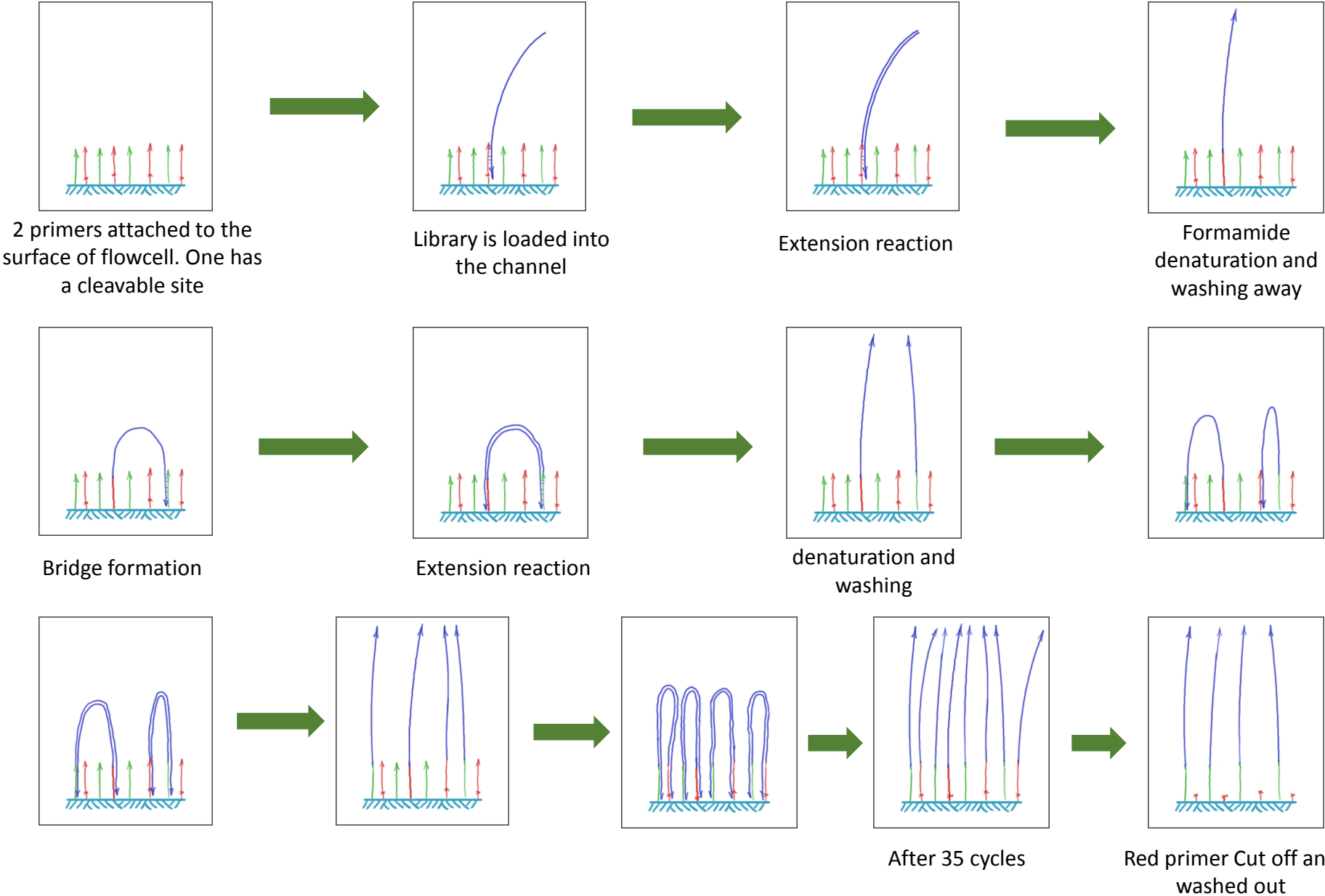


# Cluster generation

- Objective: signal amplification
- Method: bridge PCR



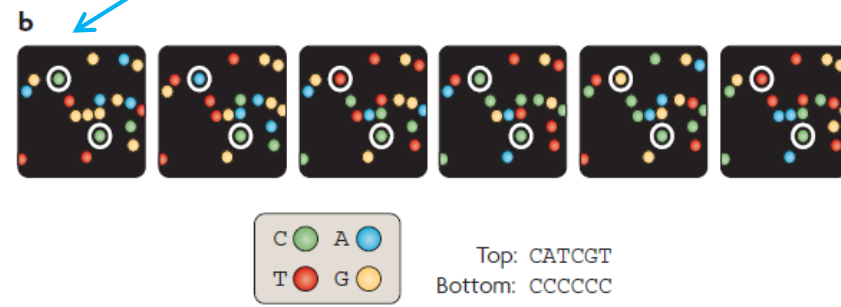
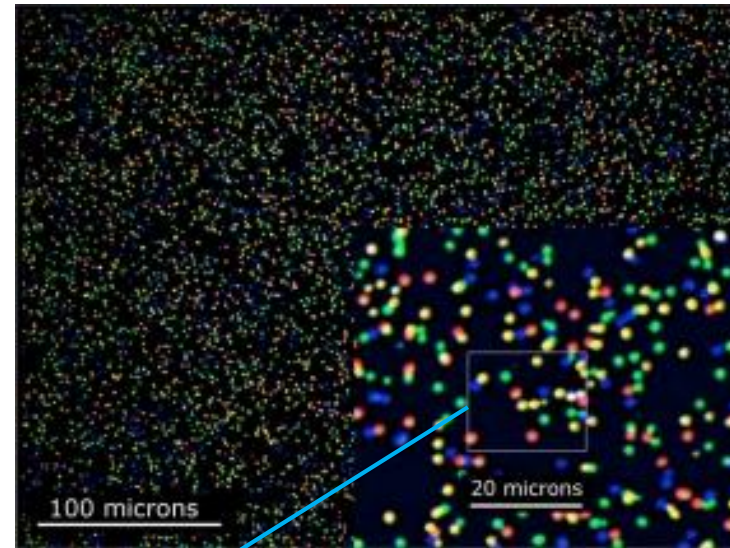
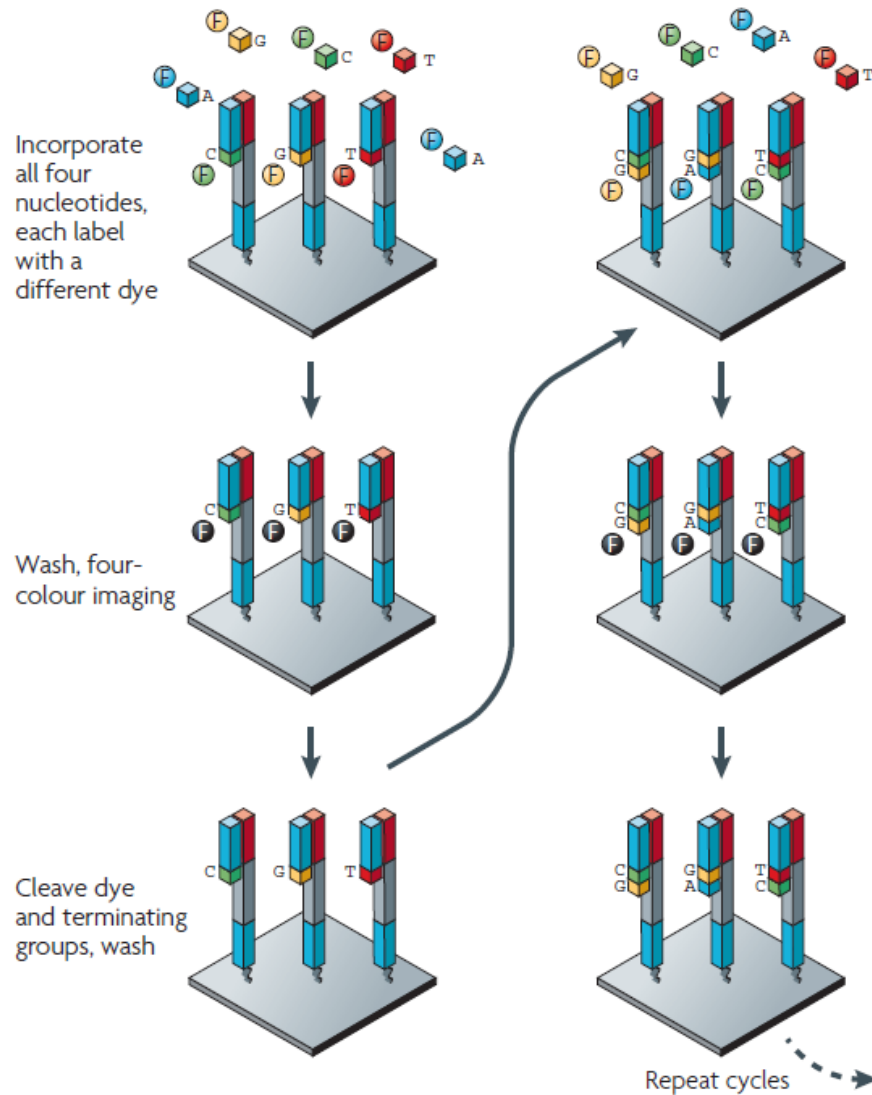
# Bridge PCR





# 3. Reversible terminator cyclic array sequencing

a Illumina/Solexa — Reversible terminators



# Output of Illumina Platform

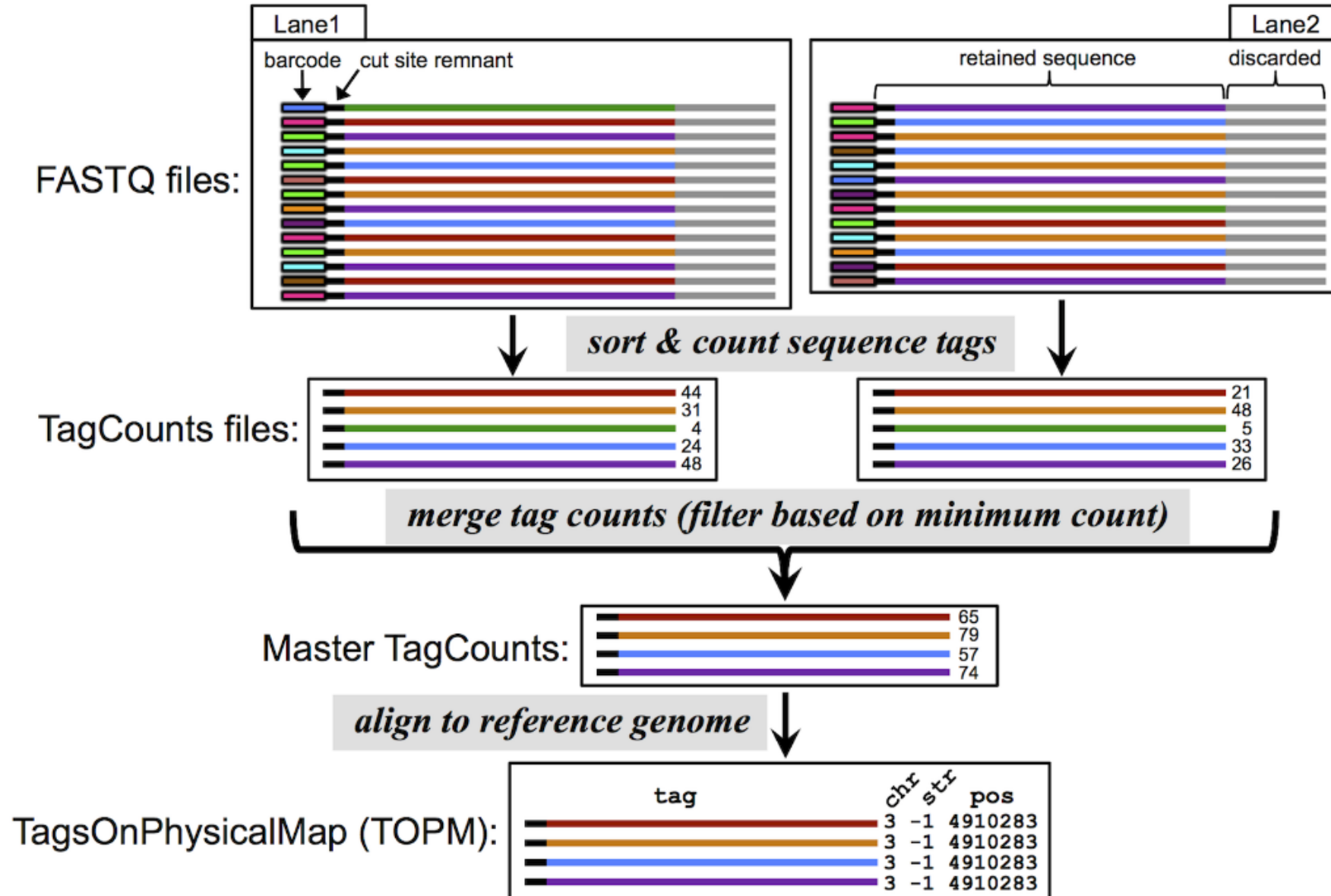
The diagram shows three callout boxes above the data table. A blue box labeled 'Barcode' points to the first three columns of the table. A red box labeled 'Cut site remnant' points to the 'CTGC' sequence in the first row. A black box labeled 'Genomic sequence' points to the rest of the sequence in the first row.

HWI-ST397	0	3	68	15896	200039	0	1	GTCGATTCTGCTGACTTCATGGCTTCTGTTGAC
HWI-ST397	0	3	68	15960	200043	0	1	GAGAATCAGCTTTTCCAACACCTTGAGTTTGAG
HWI-ST397	0	3	68	15831	200053	0	1	ATGTA CTGCACCGTTGCAAGCGAGCACCACCA
HWI-ST397	0	3	68	15867	200049	0	1	CCAGCTCAGCCTGCATTCTTTCAAAAACCTTCCA
HWI-ST397	0	3	68	15943	200048	0	1	GATTTTACTGCACATCGGTCTTGTCACACCAGC
HWI-ST397	0	3	68	15812	200062	0	1	TCACCCAGCATCACGCCCTTCACATCCAGTAA
HWI-ST397	0	3	68	15888	200067	0	1	CTTGACTGCCACCATGAATATGTGTTCCAAGTG
HWI-ST397	0	3	68	15969	200067	0	1	CCACA ACTGCTCCATCTTTCCATGAGACATTG
HWI-ST397	0	3	68	15786	200078	0	1	GTATTCTGCACACGAATCAGCTGAGACACCAAT
HWI-ST397	0	3	68	15830	200072	0	1	AATATGCCAGCAGTTAAGAGAGTTCAAGATCCA
HWI-ST397	0	3	68	15863	200073	0	1	CTCCCTGCGGGTGCGCGCGACCCATCTTCAGT
HWI-ST397	0	3	68	15762	200088	0	1	TGGTACGTCTGCGGAATGGCGTTTTTTATGCCT
HWI-ST397	0	3	68	15903	200085	0	1	GGACCTACTGCCAAGAACGGCTCACCCATCA
HWI-ST397	0	3	68	15921	200082	0	1	GAGAATCAGCGTGTACGGGGCACGGGGTGACT
HWI-ST397	0	3	68	15984	200085	0	1	TTCTCCAGCCGCATGGGCCGGAGACCAGAGAC
HWI-ST397	0	3	68	15788	200096	0	1	GCGTCAGCAAATGCCCCAACAGCCAAGTCAGC
HWI-ST397	0	3	68	15842	200099	0	1	TAGGCCATCAGCTGACTTCCCGGGTGTGGAGA
HWI-ST397	0	3	68	15876	200105	0	1	GGACCTACTGCCGCGGGACGAAAGCGGTTG
HWI-ST397	0	3	68	15937	200097	0	1	CTCCCTGTTGAAGCATGTGCAAAAGAGCTTGT
HWI-ST397	0	3	68	15958	200102	0	1	CGCCTTATCTGCCCTCGCCGGTCATGGGGAGT

64-bp sequence reads

# Data analysis

**A**



# Data analysis

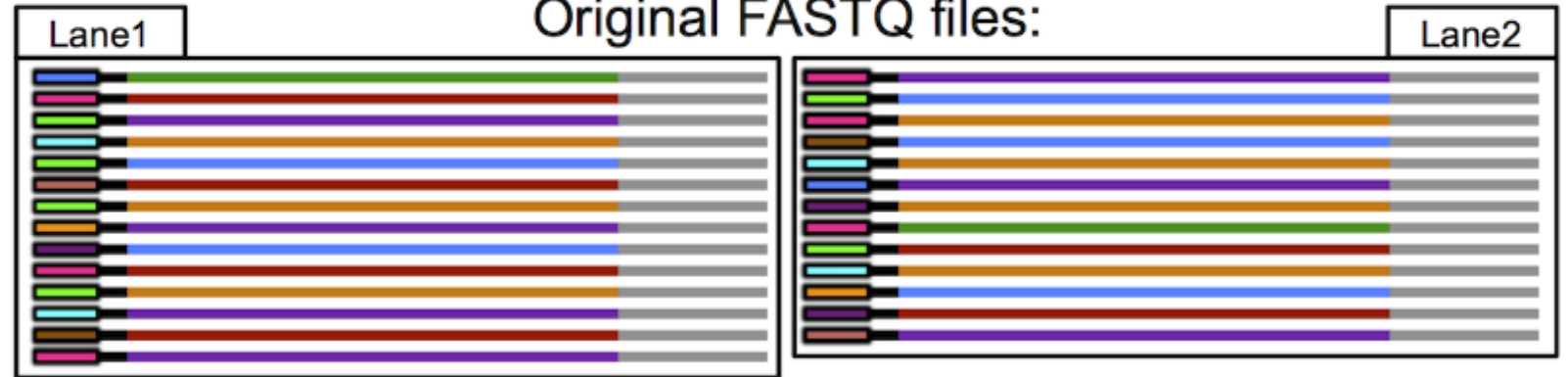
**B**

Master TagCounts:



+

Original FASTQ files:



*determine distribution of master tags among samples ("taxa")*




TagsByTaxa (TBT):

tags	taxa:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Red		0	0	1	0	0	3	0	2	1	0	0	0	1	0	2
Orange		1	0	0	3	5	0	0	0	0	4	0	0	2	0	0
Blue		0	0	3	0	0	0	0	1	0	2	0	0	0	0	7
Purple		0	6	0	0	2	0	1	0	0	0	1	3	0	0	2

# Data analysis



















## C

TagsOnPhysicalMap (TOPM):

tag	chr	str	pos
	3	-1	4910283
	3	-1	4910283
	3	-1	4910283
	3	-1	4910283





+

TagsByTaxa (TBT):

tags	taxa:																
		0	0	1	0	0	3	0	2	1	0	0	0	1	0	2	0
		1	0	0	3	5	0	0	0	0	4	0	0	2	0	0	0
		0	0	3	0	0	0	0	1	0	2	0	0	0	0	0	7
		0	6	0	0	2	0	1	0	0	0	1	3	0	0	2	0

*call and filter SNPs; update TOPM with variants*

Production-ready TOPM:

tag	chr	str	pos	var1	var2
	3	-1	4910283	15 A	36 G
	3	-1	4910283	15 C	36 G
	3	-1	4910283	15 A	36 G
	3	-1	4910283	15 A	36 T



# Autres types de marqueurs moléculaires

- Techniques principales:

- RFLP
- RAPD
- AFLP
- SSR
- SNP

- Autres techniques:

- ISSR
- CAPS
- SCAR

# Comparaison des principales techniques

Marqueur	Nombre	Codominance	Polymorphisme	Cout
Isozymes	< 90	oui	faible	faible
RFLP	illimité	oui	moyen	moyen
RAPD	illimité	non	moyen	faible
SSR	illimité	oui	très élevé	faible
AFLP	illimité	non	élevé	moyen
SNP	illimité	oui	très élevé	élevé!!!



## 2. Cartographie de QTL

# Introduction

- **Objectif:** identifier le déterminisme génétique d'un caractère
- **QTL (Quantitative trait locus):** Une région chromosomique contenant un/des gènes affectant un caractère quantitatif
- Combien de QTLs pour un caractère?
  - **Modèle de Mather:** très grand nombre de QTLs avec un effet mineur et similaire
  - **Modèle de de Robertson:** peu de QTLs avec un effet majeur et beaucoup avec un effet mineur

# Introduction

- **Cartographie de QTLs**

- Déterminer la position sur la carte génétique
- Déterminer l'effet du QTL sur la variation phénotypique du caractère ( $R^2\%$ )
- Déterminer le type d'action (additive, dominance, épistasie, combinaison)

# Principe de détection de QTLs

	QTL	FW	Marker 1	Marker 2
	qq	14	MM	AA
	qq	18	mm	aa
	qq	13	mm	AA
	QQ	21	MM	aa
	QQ	24	MM	aa
	qq	13	mm	AA
	qq	11	mm	aa
	QQ	26	mm	aa
	qq	13	MM	aa
	qq	18	mm	AA
	qq	11	mm	aa
	QQ	21	MM	AA
	qq	19	mm	aa
	QQ	20	MM	AA
	QQ	21	mm	aa
Moyenne qq/mm/aa	14.4		16.7	17.75
Moyenne QQ/MM/AA	22		18.8	17.3
Différence	7.6		2.1	-0.45

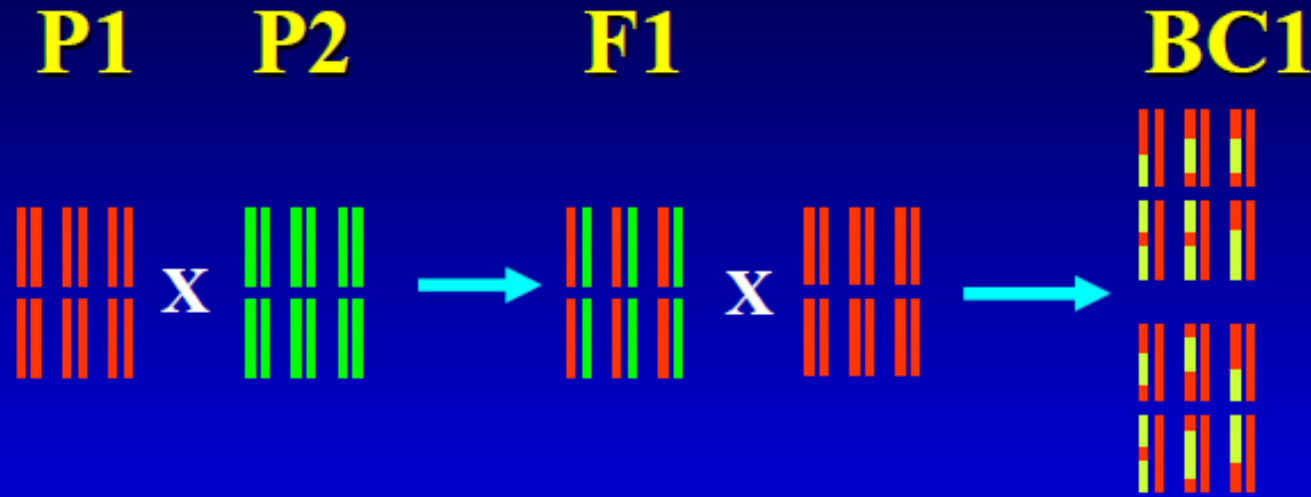
# Comment cartographier les QTLs

- Construire une population ségrégante
- Phénotyper la population pour le caractère d'intérêt
- Génotyper la population par un nombre suffisant de marqueurs
- Construire la carte génétique
- Réaliser l'analyse de QTLs (associer le génotype au phénotype)

# Populations pour cartographie de QTLs

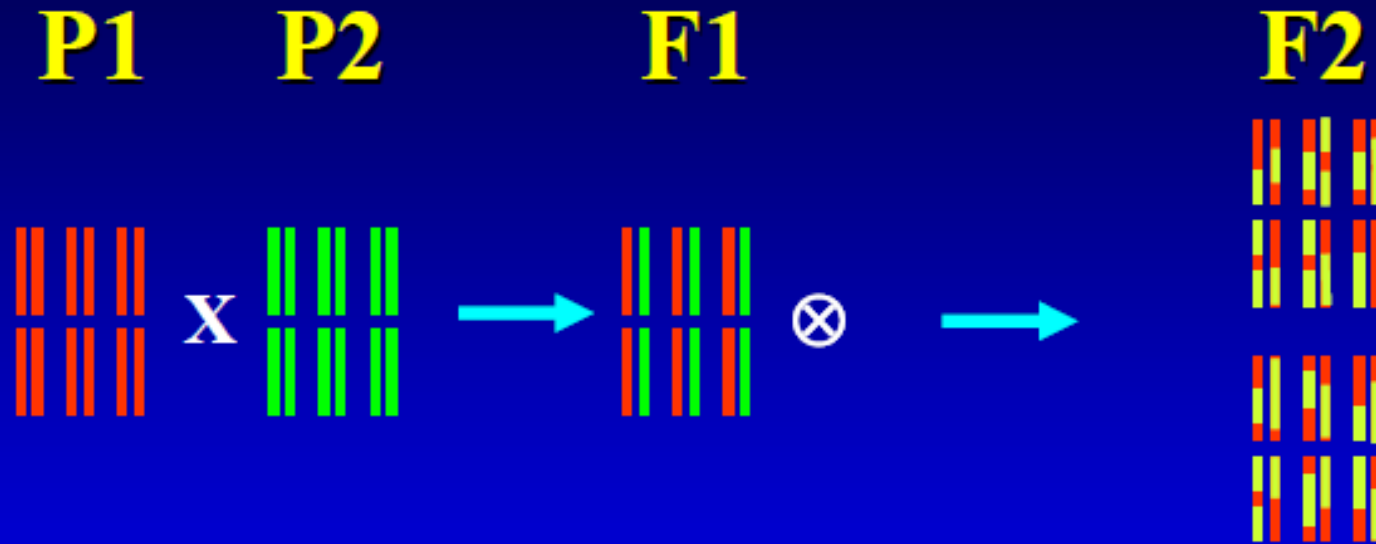
- **Espèces autogames**
  - BC1: Backcross1
  - F2
  - DHLs: Double haploid lines
  - RILs: Recombinant inbred lines
  - Autres
- **Espèces allogames**
  - F1

# Backcross



- **One meiosis.**
- **Only one of the chromosomes is recombinant**
- **Individuals are either homozygous for P1 alleles or heterozygous**

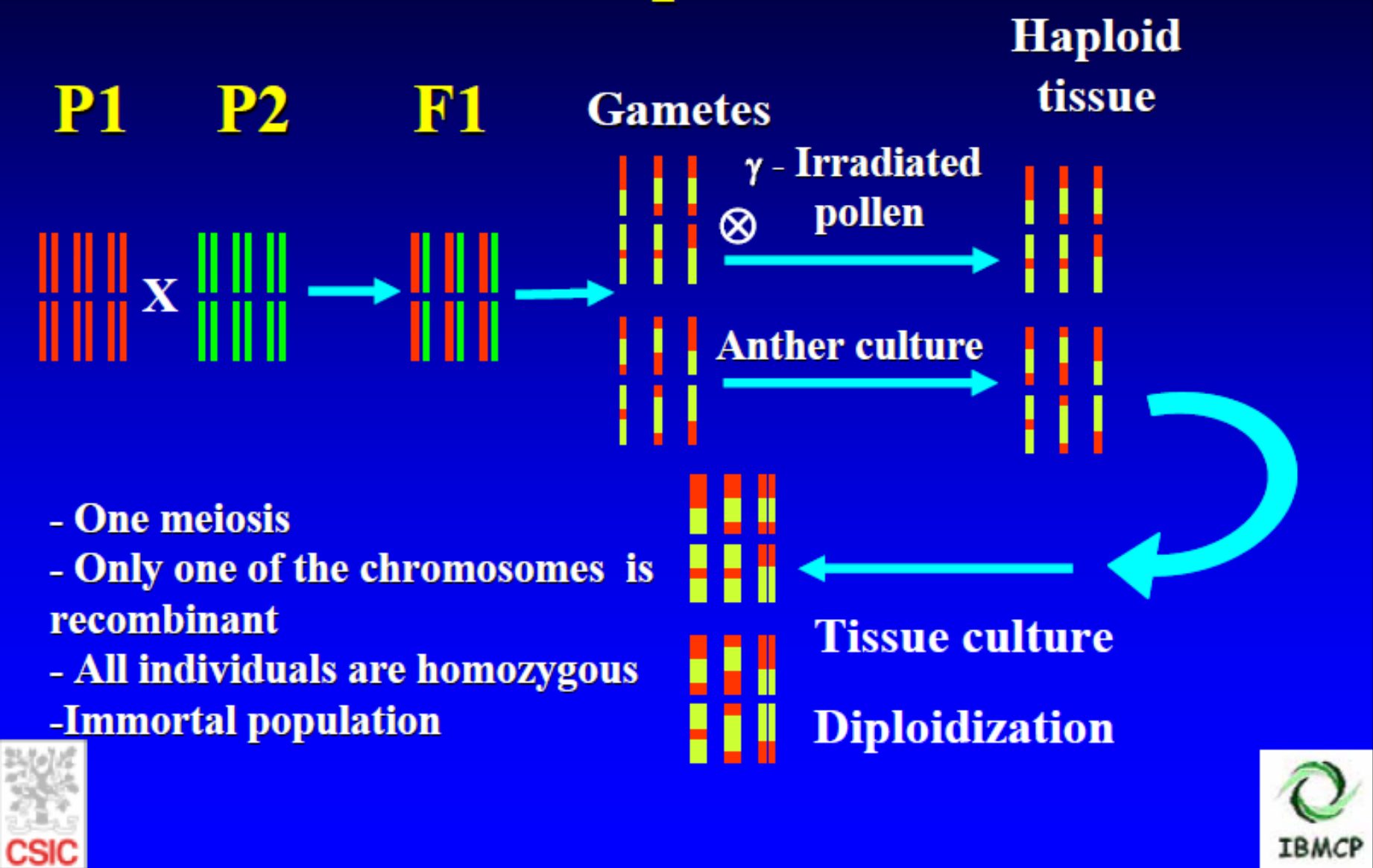
$F_2$



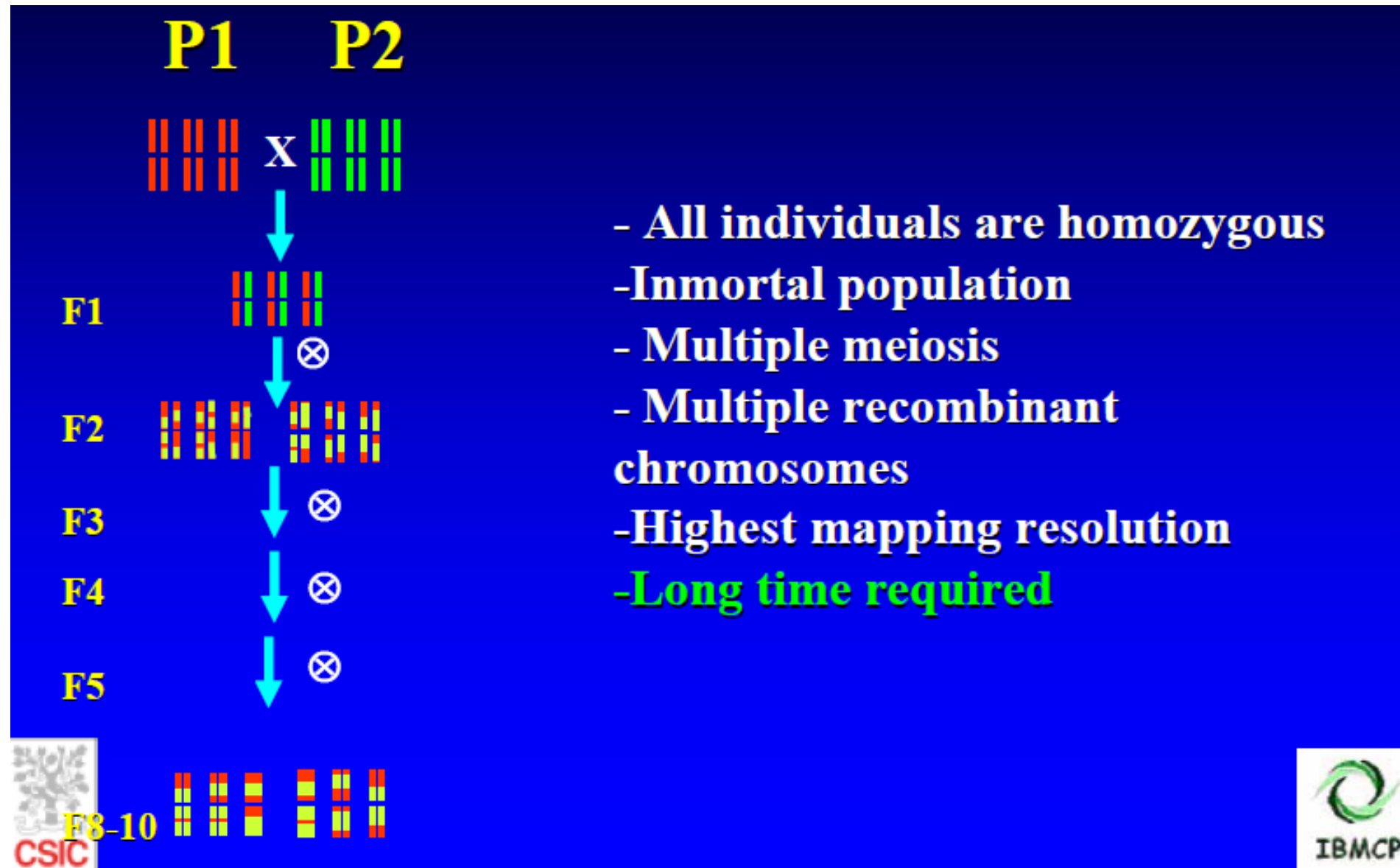
- One meiosis.
- Recombination in both mega- and microgametophyte permit better linkage resolution



# Lignées haploïdes doublées (DHL)



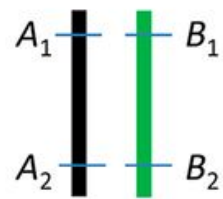
# Recombinant Inbred Lines (RILs)



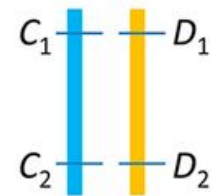
# Populations pour cartographie de QTLs

**F1**

Female parent



Male parent



×



Clonal F<sub>1</sub> Progenies



# Génotyper la population par un nombre suffisant de marqueurs

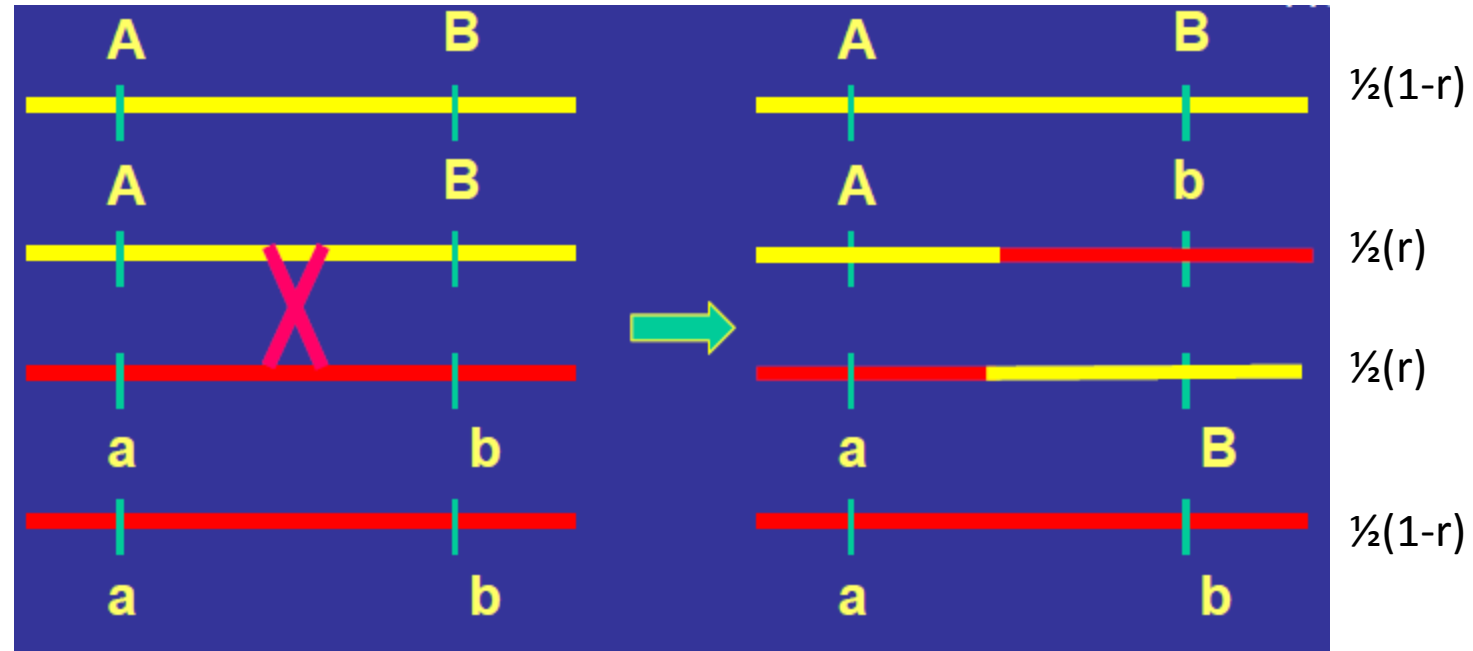
- **Quel type de marqueurs?**

- **Isozymes (isoenzymes):** enzymes présentant différentes séquence d'acides aminés mais catalysant la même réaction chimique
- **RFLPs:** Restriction Fragment Length Polymorphism
- **RAPDs:** Random Amplified Polymorphic DNA
- **AFLPs:** Amplified Fragment Length Polymorphism
- **SSRs:** Simple Sequence Repeat
- **SNPs:** Single Nucleotide Polymorphism

# Construction de cartes génétiques

- **Objectif:** déterminer la position relative des marqueurs sur les groupe de liaison (linkage groups, chromosomes)
- **Principe:** la distance entre les marqueurs est proportionnelle à la fréquence de recombinaison entre eux

# Construction de cartes génétiques



Fréquence de recombinaison ( $r$ )= nombre de recombinants/total

$$0 < r < 0.5$$

# Construction de cartes génétiques

**r is not the optimal parameter for map construction**

- r tends to a maximum of 0.5
- r is not additive
- double recombinations are not considered

**INTERFERENCE:** When a crossover occurs, it is less probable that another one will occur at the same region of the chromosome.

# Fonction de calcul de la distance génétique

**Haldane: assumes no interference**

**Kosambi: includes some degree of interference**

**Distance unit: CENTIMORGAN (cM)**

**1 cM  $\approx$  1 % recomb**



# Fonction de calcul de la distance génétique

**Haldane:**  $\begin{cases} r = \frac{1}{2} (1 - e^{-2x}) \\ x = -\frac{1}{2} \ln(1 - 2r) \end{cases}$

r =	0	0.10	0.30	0.32	0.40	0.43	0.491
x =	0	11	46	50	81	100	200

rec. freq.  
cM

**Kosambi:**  $\begin{cases} x = \frac{1}{4} \ln\left(\frac{1 + 2r}{1 - 2r}\right) \\ r = \frac{1}{2} \tanh(2x) \end{cases}$

r =	0	0.10	0.30	0.38	0.40	0.48	0.4997
x =	0	10	35	50	55	100	200

rec. freq.  
cM

# Le test statistique

**LOD score: Logarithm of odds**

$$\text{LOD} = \text{Log}_{10} \left( \frac{\text{Likelihood when } r = \hat{r}}{\text{Likelihood when } r = 0.5} \right)$$

**LOD values > 3.0 (1.000 times more probable that the data indicate linkage than independence) are usually accepted as a threshold of significance**

# Comment construire une carte?

- Calculer les distances entre tout les paires de marqueurs
- Etablir les groupes de liaison: grouper ensemble tout les marqueurs ayant des liaisons significatives (en fonction du seuil fixé pour LOD)
- Ordonner les marqueurs à l'intérieur des groupes
- MapMaker, IciMapping, JoinMap (payant)

# Méthodes d'analyse de QTLs

- **Single marker analysis**
  - Study independently each marker
- **Interval Mapping**
  - Study the presence of a QTL within the interval defined between two linked markers
- **Composite Interval Mapping**
  - Use markers on other chromosomes as cofactors to increase the resolution

Qgene, QTL cartographer, R/QTL, IciMapping