1-i) Review on Molecular Marker: A tool for genetic diversity research

Prof. K. Watanabe

Gene Research Center,

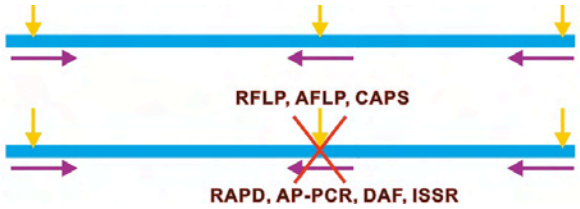
Univ. Tsukuba, Japan

Introduction

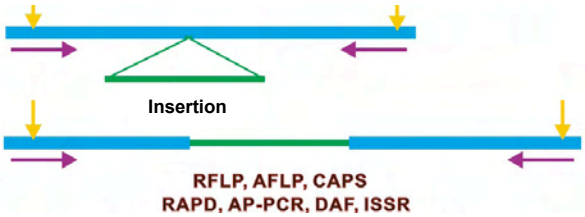
- The development of genetic marker
  - Morphologic marker (eg. flower color, plant height etc.)
  - Protein marker / Biochemical marker (eg. isozyme)
  - DNA marker / Molecular marker (RFLP, RAPD, SSR etc.)
- Molecular nature of naturally occurred polymorphism
  - Point mutation
  - Insertion / deletion
  - DNA rearrangement

The molecular basic of DNA marker

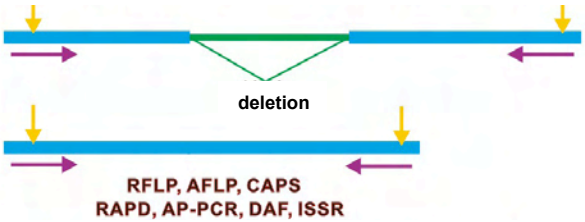
1. Point mutation between restriction sites (PCR primer binding sites)



2. Insertion between restriction sites (PCR primer binding sites)



3. Deletion between restriction sites (PCR primer binding sites)



4. Number of tandem repeats varying between restriction sites (PCR primer binding sites)

5. Single nucleotide mutation

Introduction

- Some regions of genome are significantly more polymorphic than singly copy sequences
  - Tandem repeats
- Synteny
  - In the use of molecular marker, an important observation is the finding that many distantly related species have co-linear maps for portions of their genomes.
  - Solanaceae
  - Gramineae
- Locus & allele
- Allele frequency & heterozygosity
- Dominant & co-dominant
- Polymorphism information

Classification of Molecular Marker by Detection Technology

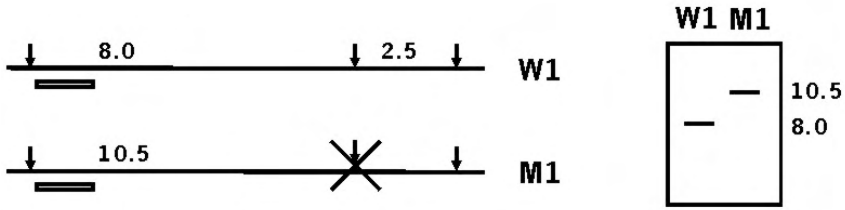
- Based on DNA-DNA hybridization
- Based on PCR technology
- Based on restriction digest and PCR
- Based on DNA sequencing and microarray

Based on DNA-DNA hybridization

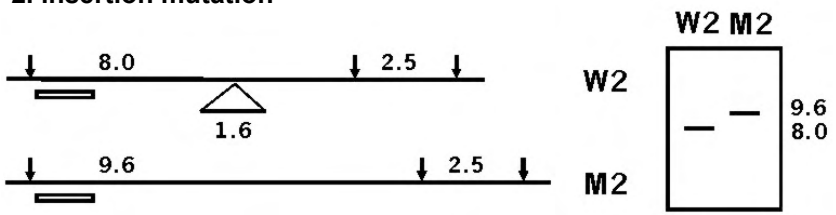
- RFLP, restriction fragment length polymorphism
- VNTR, variable number of tandem repeats

RFLP

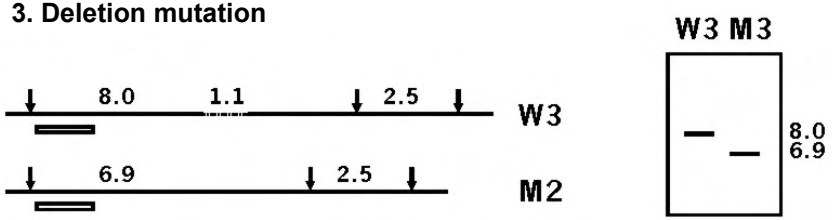
1. mutation in restriction site



2. insertion mutation



3. Deletion mutation



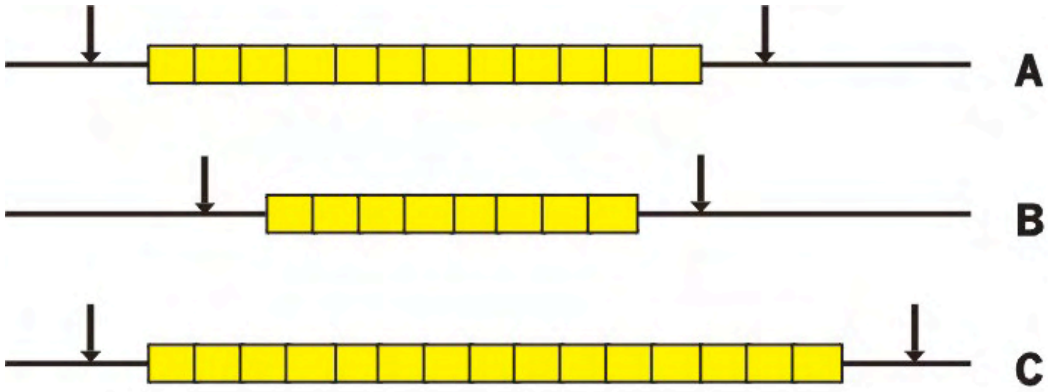
: **restriction site**

: **probe**

W: Wild type

M: Mutant

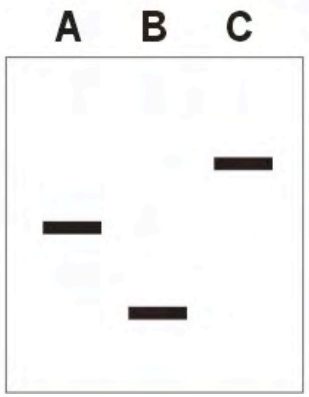
VNTR

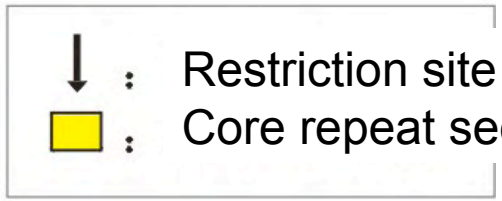


Restriction digest

Hybridization with tandem repeats sequence as probe

autoradiography



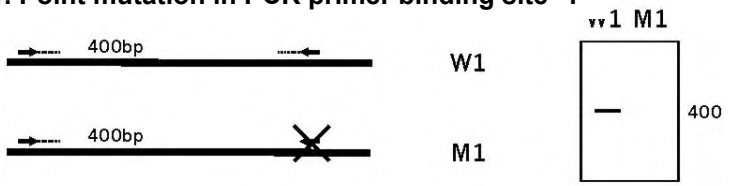


Based on PCR technology

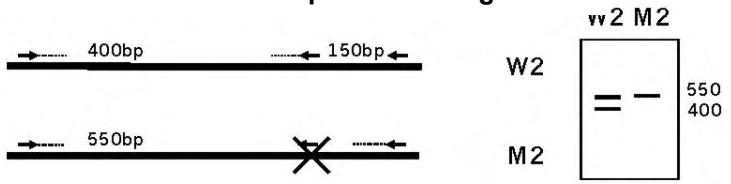
- Based on random primers
  - RAPD, random amplified polymorphismic DNA
  - AP-PCR, arbitrarily primed PCR
  - DAF, DNA amplification fingerprinting
  - ISSR, inter-simple sequence repeats
- Based on special primers
  - SSR, simple sequence repeats
  - SCAR, sequence characterized amplified region
  - STS, sequence-tagged site
  - RGA, resistance gene analogs

RAPD

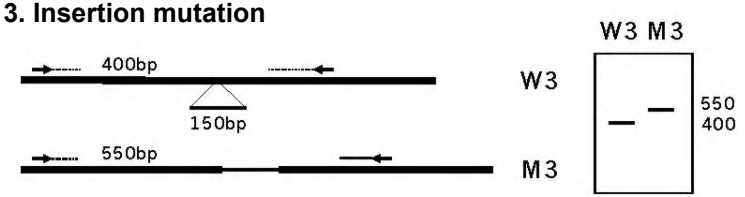
1. Point mutation in PCR primer binding site -1



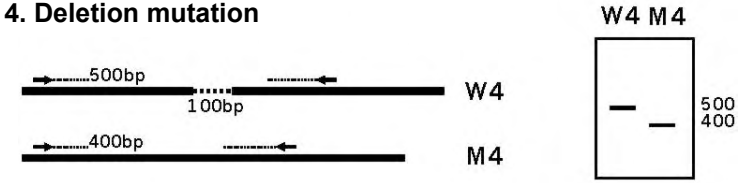
2. Point mutation in PCR primer binding site -2



3. Insertion mutation



4. Deletion mutation

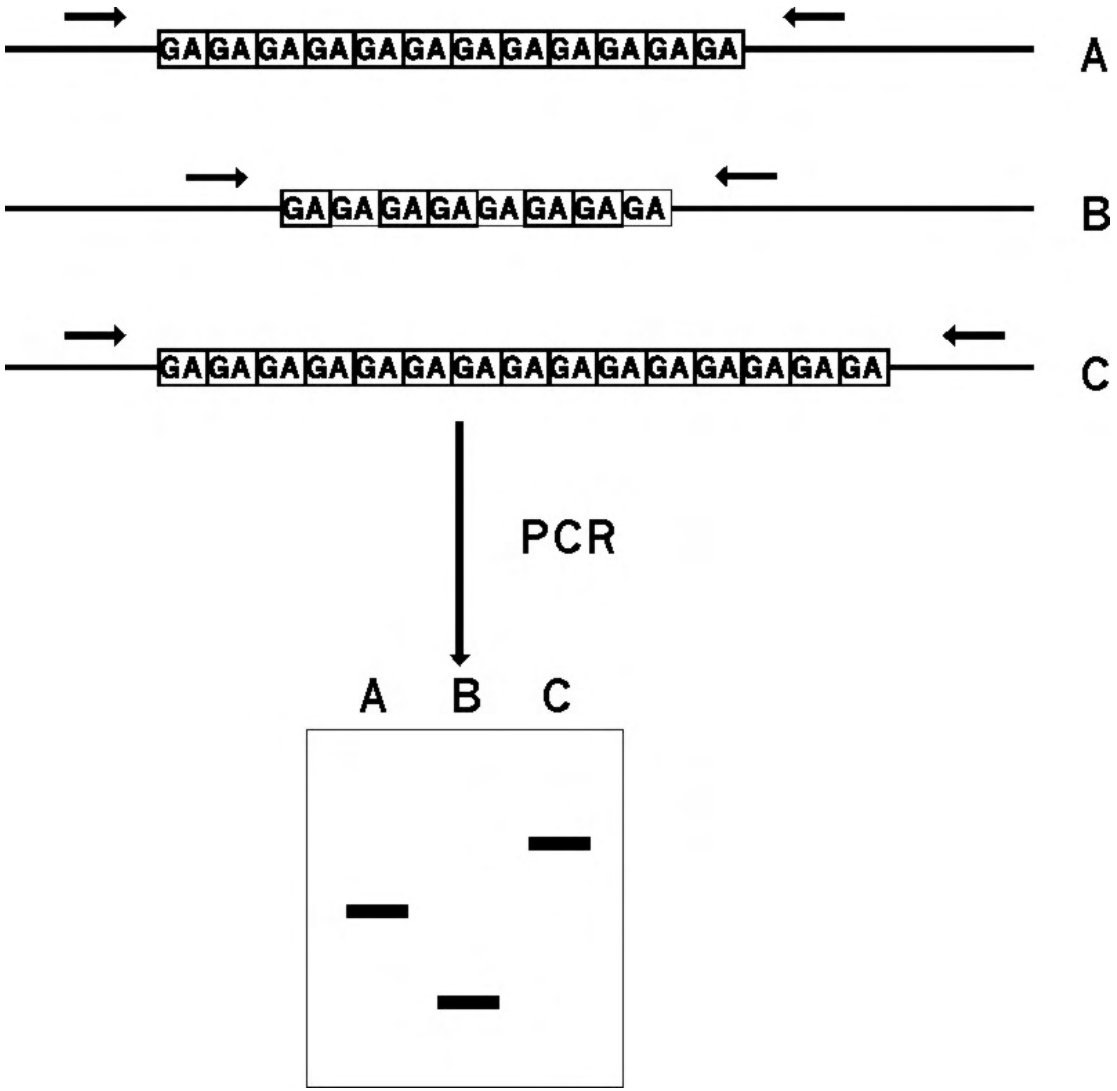


→ : **primer**

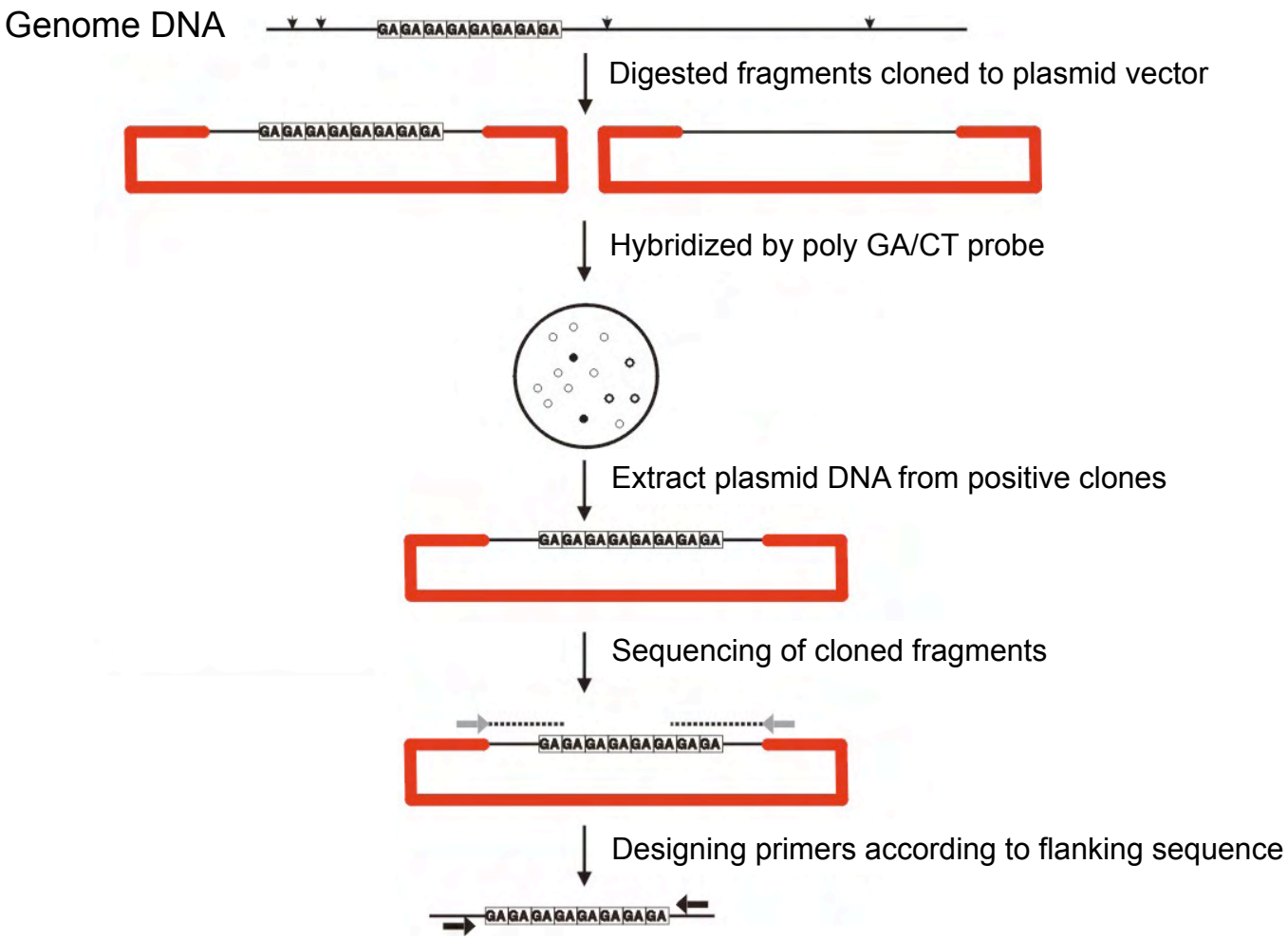
W: **Wild type**

M: **Mutant**

SSR


GAGAGAGAGAGAGAGAGAGAGAGAGA
A
GAGAGAGAGAGAGAGA
B
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
C
PCR
A
B
C

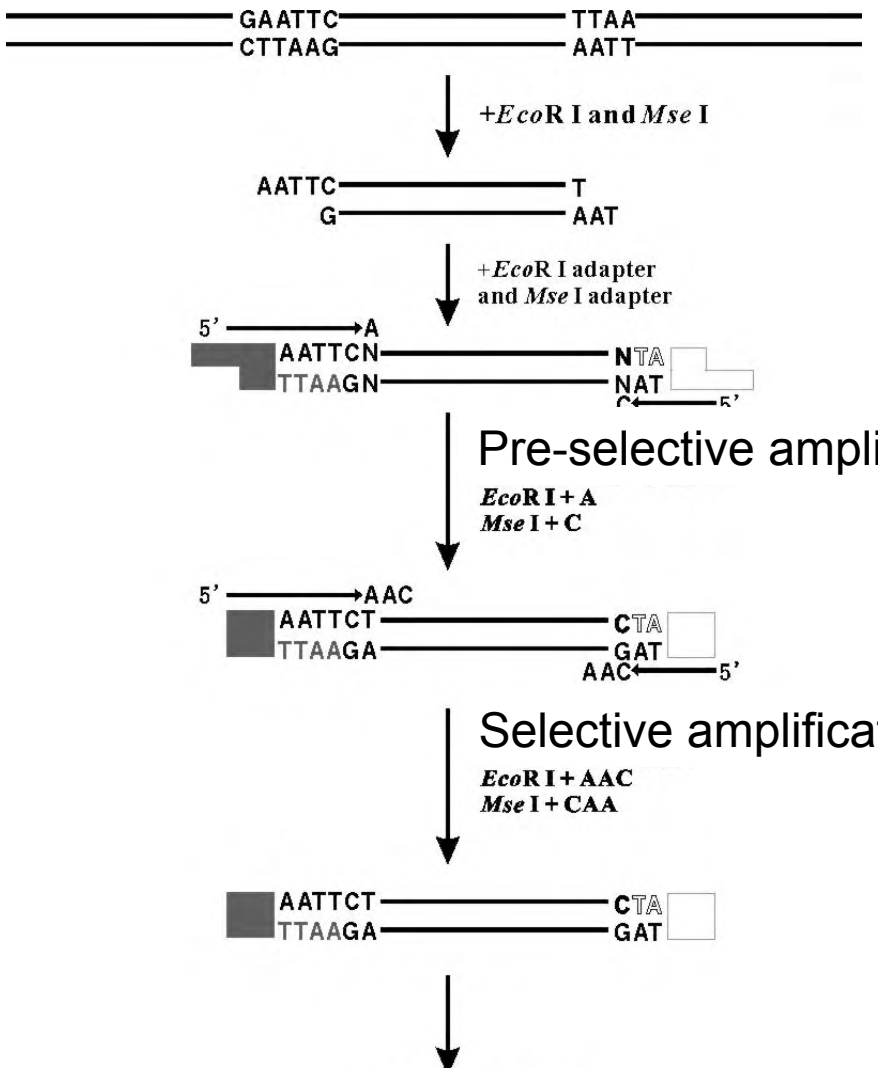
Developing SSR Primers



Based on restriction digest and PCR

- AFLP, amplified fragment length polymorphism
- CAPS, cleaved amplified polymorphic sequence

Procedure of AFLP



Denatured Gel Electrophoresis

Based on DNA sequencing and microarray

- SNP, single nucleotide polymorphism
  - SSCP (Single-strand conformation polymorphism)
  - DGGE (Denaturing gradient gel electrophoresis)
  - ASA (Allele-specific amplification)
  - GBA (Genetic bit analysis)
  - Oligonucleotide chip-based hybridization
  - MALDI-TOF MS (Matrix assisted laser desorption ionization, time of flight mass spectrometry)

Application of Molecular Marker

- Phylogeny
- Genetic diversity
- Molecular Mapping
- Gene tagging
- MAS, marker assisted selection
- Genebank management: duplicate identification
- Fingerprinting
- Quality testing

Phylogenetics

- **Phylogenetics** is the taxonomical classification of organisms based on how closely they are related in terms of evolutionary differences
- A **phylogeny** (or **phylogenesis**) is the origin and evolution of a set of organisms, usually species . A major task of systematics is to determine the ancestral relationships among known species (both living and extinct), and the most commonly used methods to infer phylogenies include cladistics and phenetics .

http://research.amnh.org/users/siddall/methods/

Phylogenetics: *just* methods

The number of available methods in phylogeny and systematics may appear to be daunting not only to beginning students but also to well-established researchers. Admittedly, anyone with a copy of, say, PAUP* can learn how to "point and shoot" to make a neighbor-joining tree, or a cladogram or a likelihood tree, but this doesn't really mean you know what you're doing and we suspect that you'd probably like to know. This site is designed to do that. To provide you with an understanding of the various methods you might encounter so as to allow you to make informed decisions about what you think is defensible and appropriate.

TOPICS

1. Unweighted Pair Group Method (UPGMA, phenetics)
   Weighted Pair Group Method (WPGMA, phenetics)
2. Hennigian Argumentation (cladistics)
   Constructing Wagner Trees (cladistics)
3. Parsimony Analysis - optimization, tree searching (cladistics)
4. implementations in software
5. weighting
6. Distance Analyses ('minimum evolution') and models of stochastic DNA change
7. Neighbor Joining
8. Maximum Likelihood
9. Differences of opinion between parsimony and likelihood
10. Multiple Sequence Alignment

    Optimization Alignment - Optalign, Poy
    Character Bootstrapping
    Character Jackknifing
    Taxonomic Jackknifing
    Noise Analysis
    Bremer Support
    Cospeciation (e.g., host-parasite) Analysis
    Gene-tree/Species-tree Analysis
    Stratigraphic Analysis
    Biogeographic Analysis

Contact: Mark E. Siddall msiddall@umich.edu http://research.amnh.org/~siddall/ ph: 212 769 5638

Molecular Mapping

- Populations (parents & size)
  - F_2
  - BC_1, back cross
  - RI, recombinant inbred
  - DH, double haploid

Comparative of average physical distance and locus distance in different organisms

Species	Genome size (kb)	Genetic distance (cM)	kb / cM
Phage T_4	1.6×10^2	800	0.2
E. coli	4.2×10^3	1,750	2.4
Yeast	2.0×10^4	4,200	4.8
Fungus	2.7×10^4	1,000	27.0
Nematode	8.0×10^4	320	250.0
Drosophila	1.4×10^5	280	500.0
Rice	4.5×10^5	1,500	300.0
Mouse	3.0×10^6	1,700	1,800.0
Human race	3.3×10^6	3,300	1,000.0
Maize	2.5×10^6	2,500	1,000.0

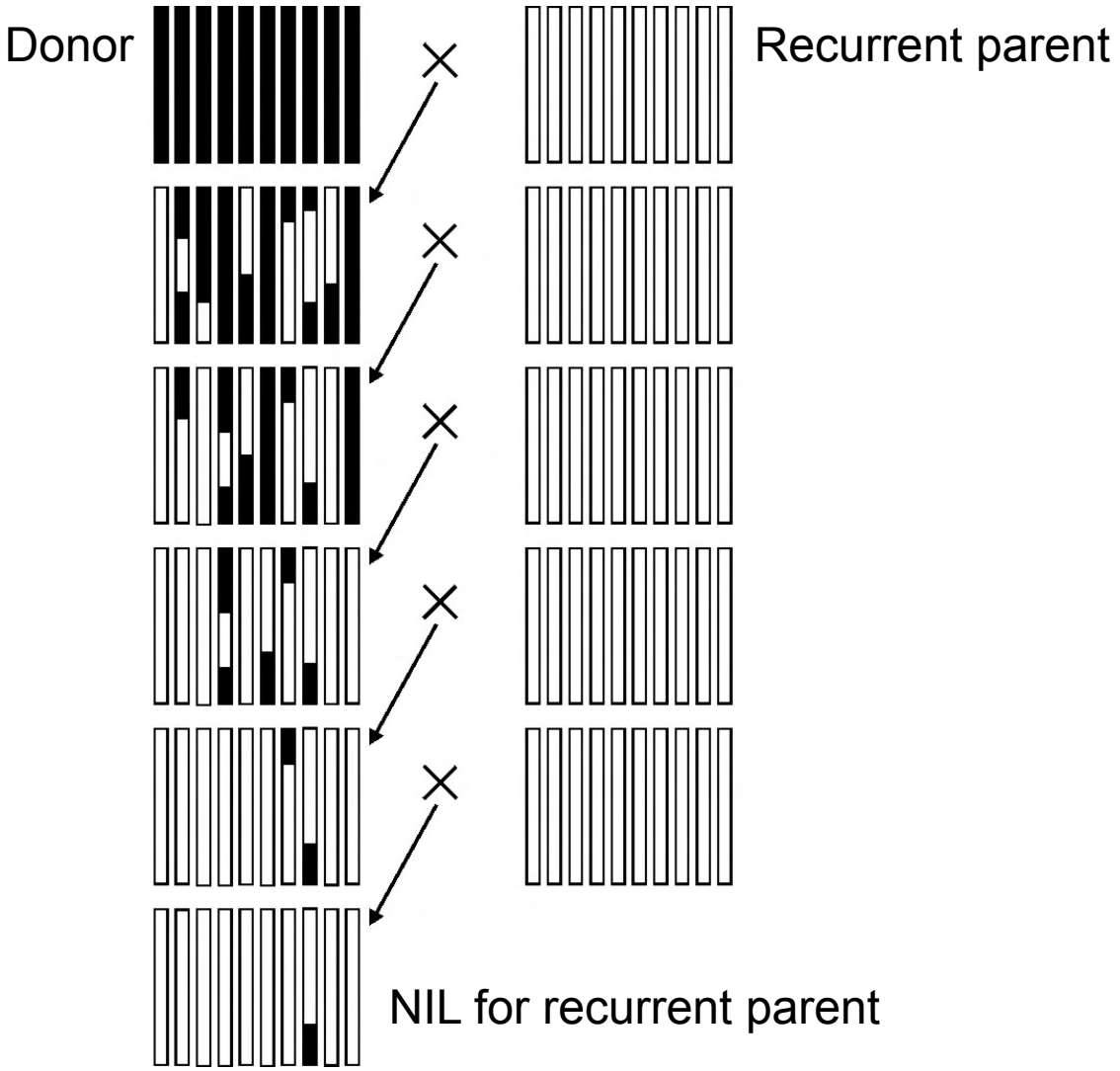
Needed marker number to reach specific saturated genetic map

Species		Human race	Rice	Maize	Arabidopsis	Tomato
Genome size	(kb)	3.3×10^6	4.5×10^5	2.5×10^6	7.0×10^4	7.1×10^5
	(cM)	3300	1500	2500	500	1500
	kb/cM	1000	300	1000	140	473
Map saturation	20cM	165	75	125	25	75
	10cM	330	150	250	50	150
	5cM	660	300	500	100	300
	1cM	3300	1500	2500	500	1500
	0.5cM	6600	3000	5000	1000	3000

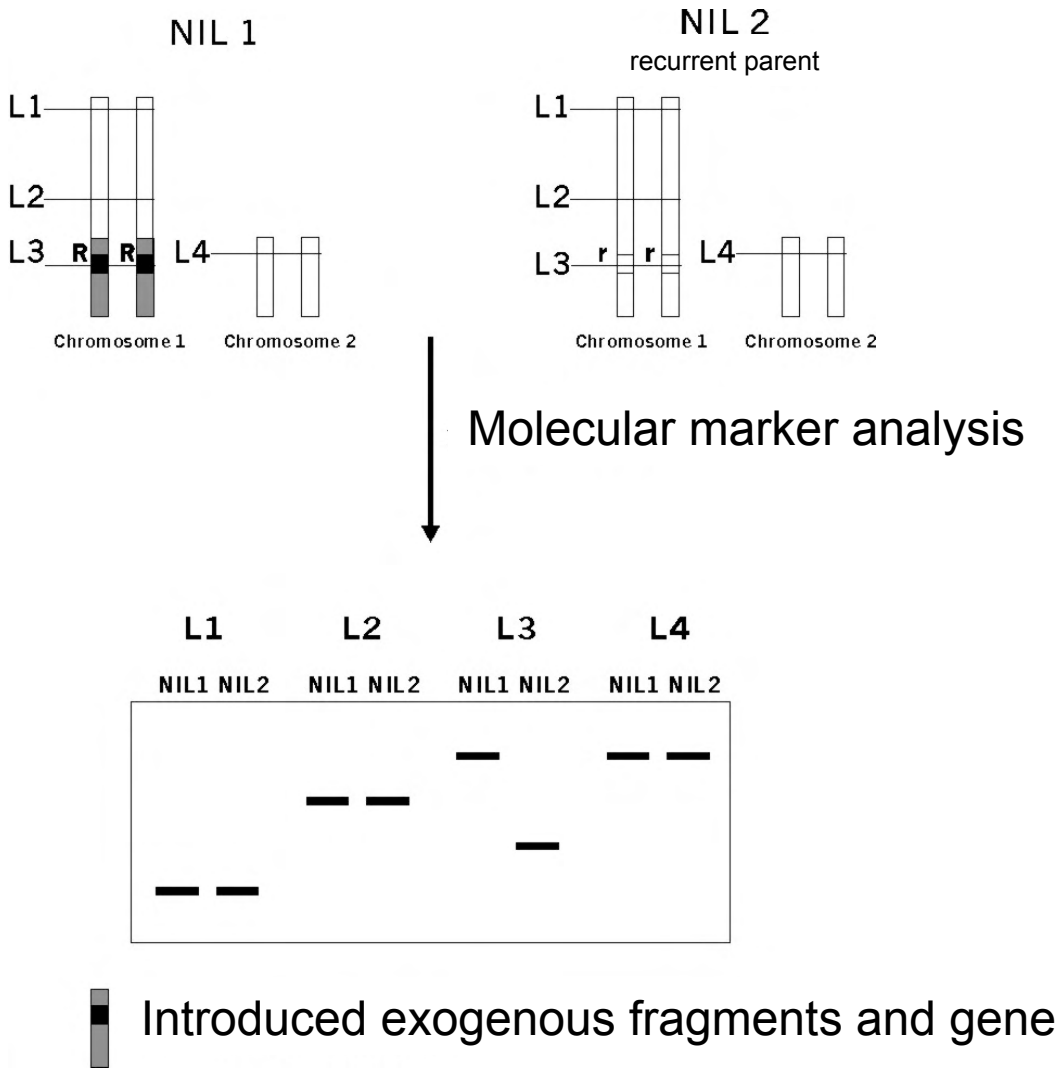
Gene Tagging for Qualitative Trait

- NIL, near isogenic lines
- BSA, bulked segregant analysis

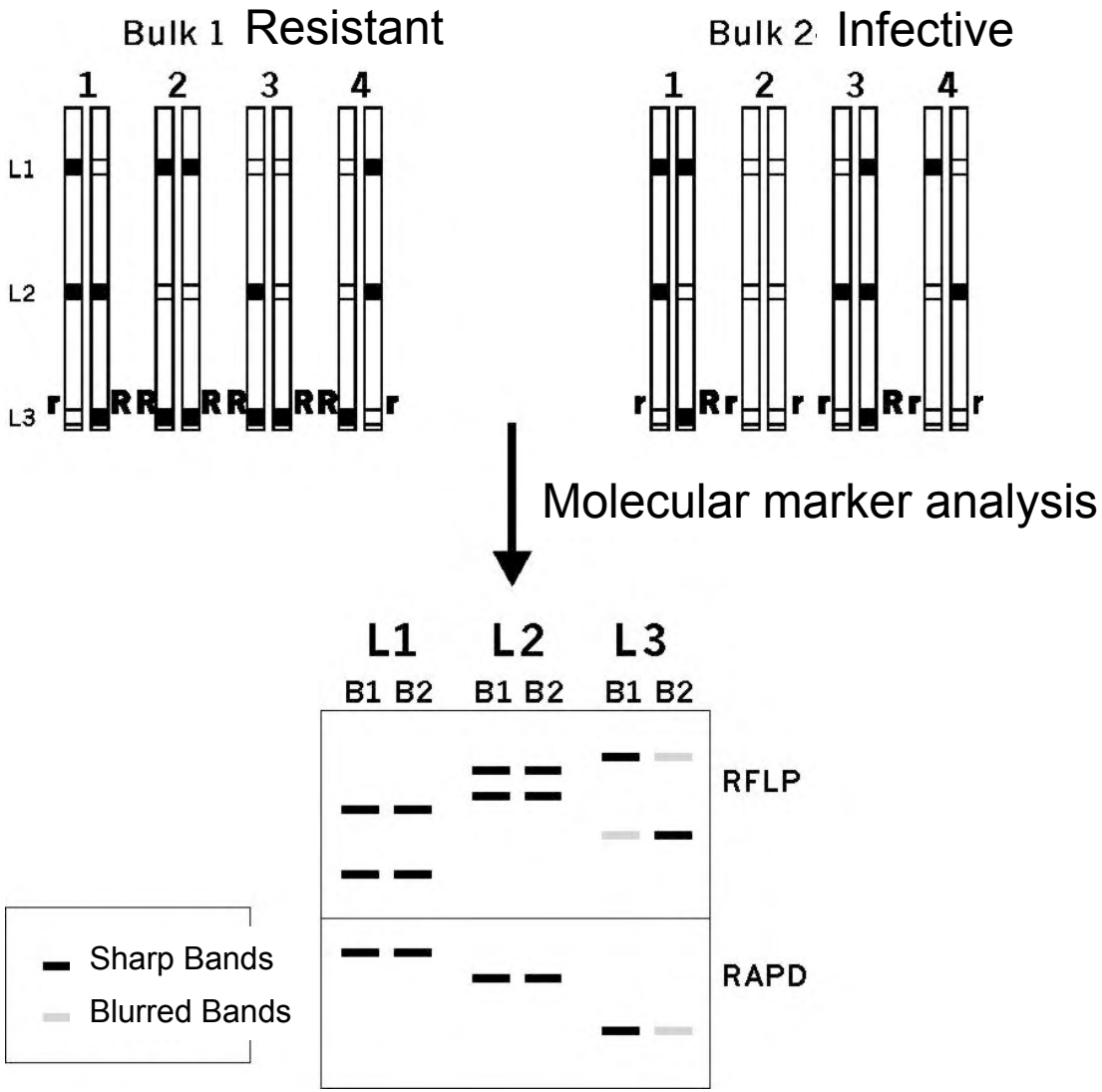
Diagram of developing NIL



Gene Tagging by NIL



Gene Tagging by BSA



QTL Mapping

- QTL, Quantitative Trait Loci
  - mono-marker method
  - interval mapping, IM
  - composite interval mapping, CIM

MAS in modern breeding

- Foreground selection
- Background selection
- Gene pyramiding
- Gene transgression

Thanks