Mapping and Mapping Populations

Types of mapping populations

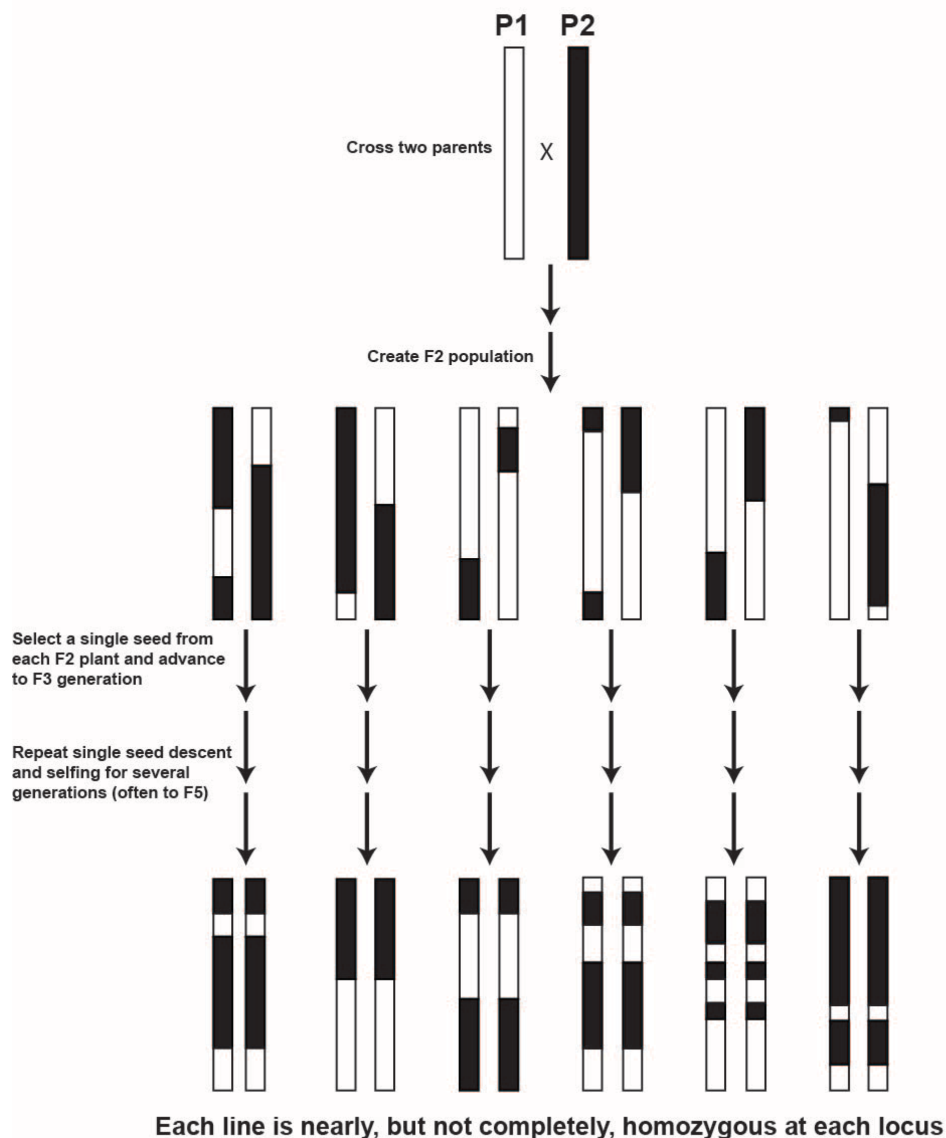
- F_2
  - Two F_1 individuals are intermated
- Backcross
  - Cross of a recurrent parent to a F_1
- Recombinant Inbred Lines (RILs; F_2-derived lines)
  - Developed by single seed descent through multiple generations of selfing

Homozygosity of Recombinant Inbred Lines

RI Population Degree of Inbreeding	% within-line homozygosity at each locus
$F_{2:3}$	75.0
$F_{2:4}$	87.5
$F_{2:5}$	92.25
$F_{2:6}$	96.875
$F_{2:7}$	98.4375
$F_{2:8}$	99.21875

Value of Recombinant Inbred Populations

- Eternal population
- Phenotypic data can be collected from replicated trials to ensure accuracy
- Large field trials can be performed for major production traits
- Only problem
    - Dominance and epistasis cannot be measured because no heterozygotes are available



Segregation Ratios for Mapping Populations

Population	**Codominant marker**	**Dominant marker**
F_2	1:2:1	3:1
Backcross	1:1	1:1*
Recombinant inbred	1:1	1:1

***** To score a dominant marker in a backcross population, you must cross the recessive parent with the F1 plant. Therefore to score a dominant marker (such as a RAPD), you would need to create two populations, each one developed by backcrossing to one of the two parents. For this reason, backcross populations have not been used for mapping dominant markers.

One of the first molecular mapping papers (Bernatzky and Tanksley. 1986. Genetics 112:887)

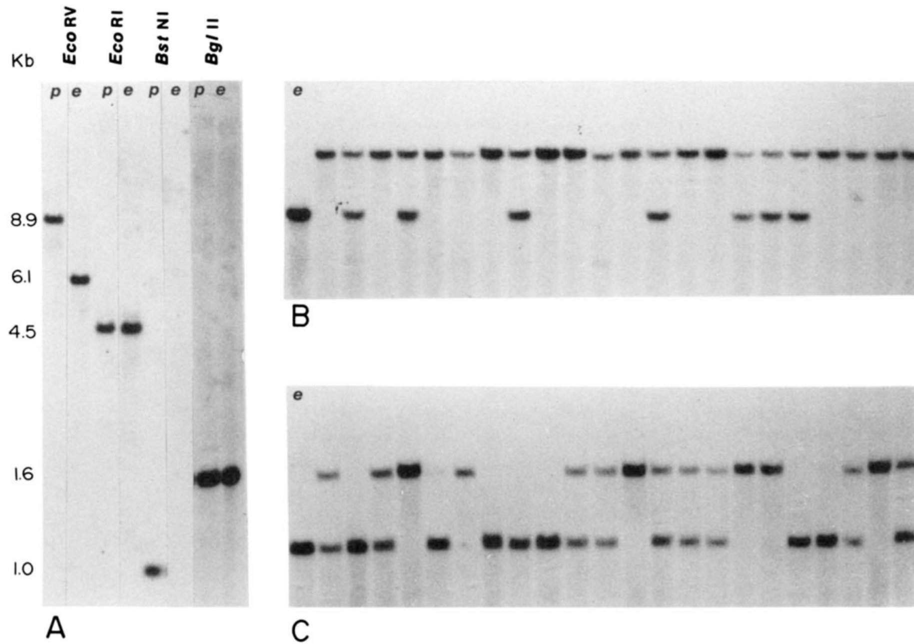


FIGURE 2.—A, Restriction enzyme survey of *L. pennellii* (p) and *L. esculentum* (e) probed with clone 3-41 (*CD14*). The values at left are the fragment sizes in kilobases. B, Backcross progeny DNA (*L. pennellii* as the recurrent parent) digested with *Eco*RV and probed with 3-41. C, F_2 progeny DNA digested with *Eco*RV and probed with 3-41.

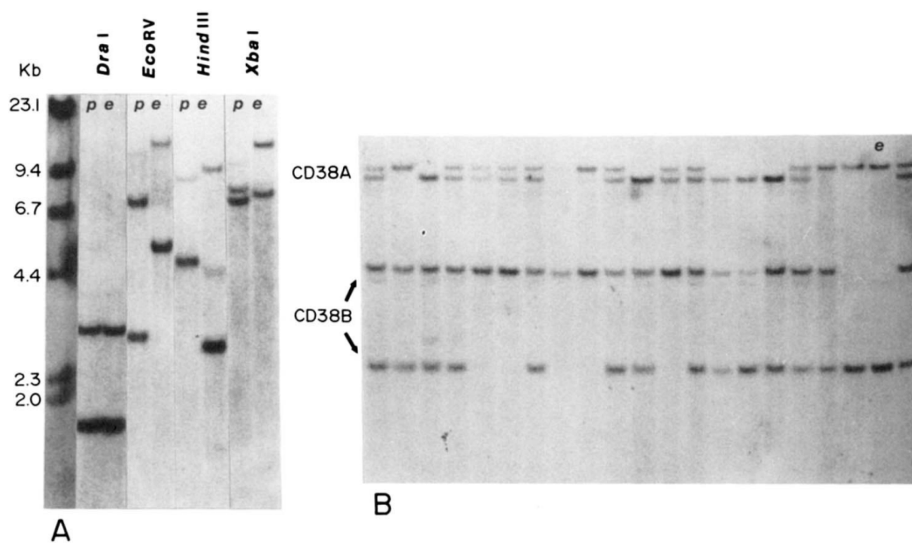
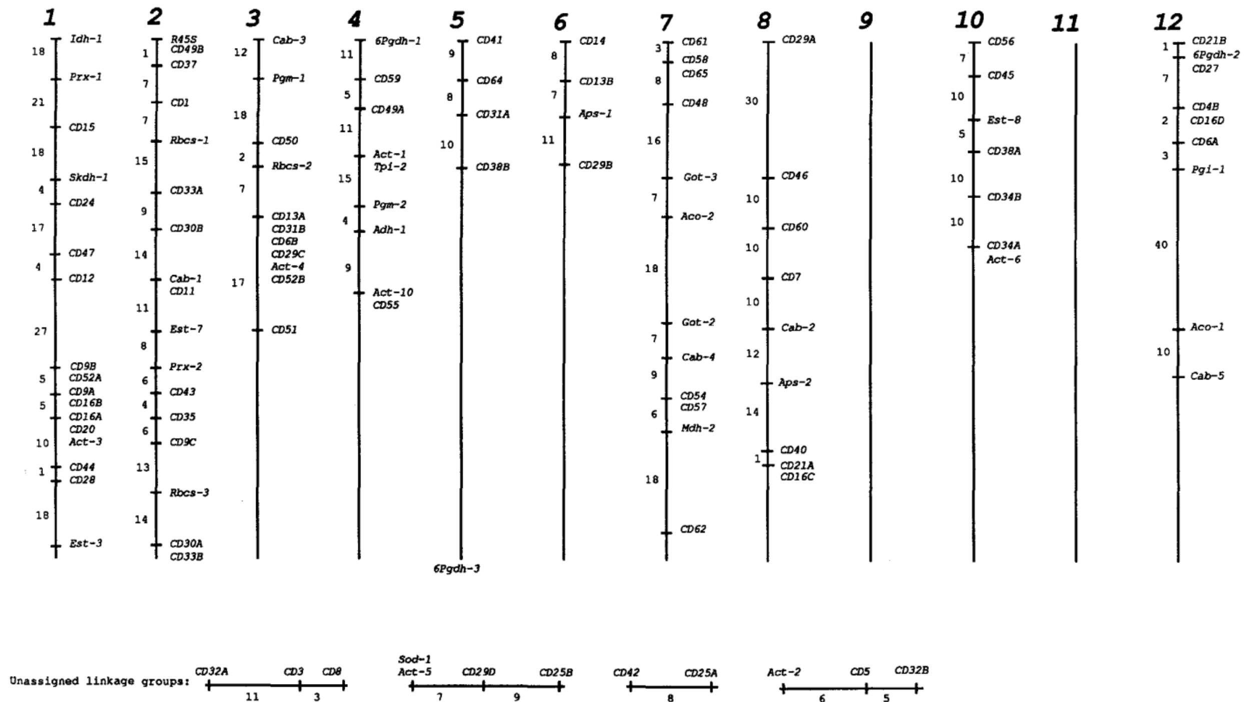


FIGURE 3.—A, Restriction enzyme survey of *L. pennellii* (p) and *L. esculentum* (e) probed with clone 3-275 (*CD38A* and *B*). The first lane is DNA digested with *Hin*dIII, and the fragment sizes are indicated at left. B, F_2 progeny DNA digested with *Hin*dIII and probed with 3-275.

One of the first molecular marker maps for plants (tomato). (Bernatzky and Tanksley. 1986. Genetics 112:887)



A first plant molecular marker map

- cDNA map of tomato
- Notice the limited number of markers
- Isozyme markers also included

Arabidopsis RAPD map (Reiter et al. 1992. Proc. Natl. Acad. Sci., USA 89:1477**)**

Segregating population gel

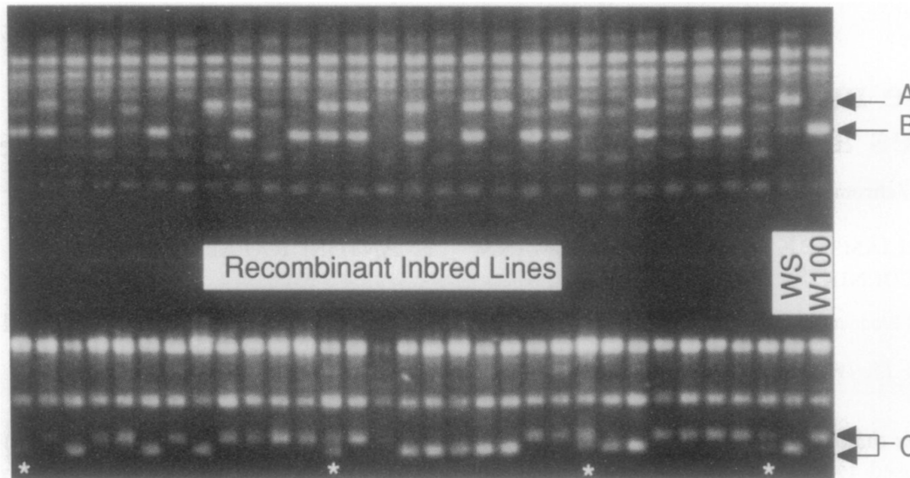


FIG. 1. Example of RAPD band segregation in RI populations. Ethidium bromide-stained electrophoretic pattern of DNA amplified from individual RI or parental lines. Arrows indicate two independent dominant markers amplified by primer r239 (A and B) and a single codominant marker amplified by primer rap14e (C). Asterisks indicate heterozygotes identified by primer rap14e.

Examples

- **(A and B)** Two dominant loci segregating
- **(C)** One dominant locus segregating

Linkage map

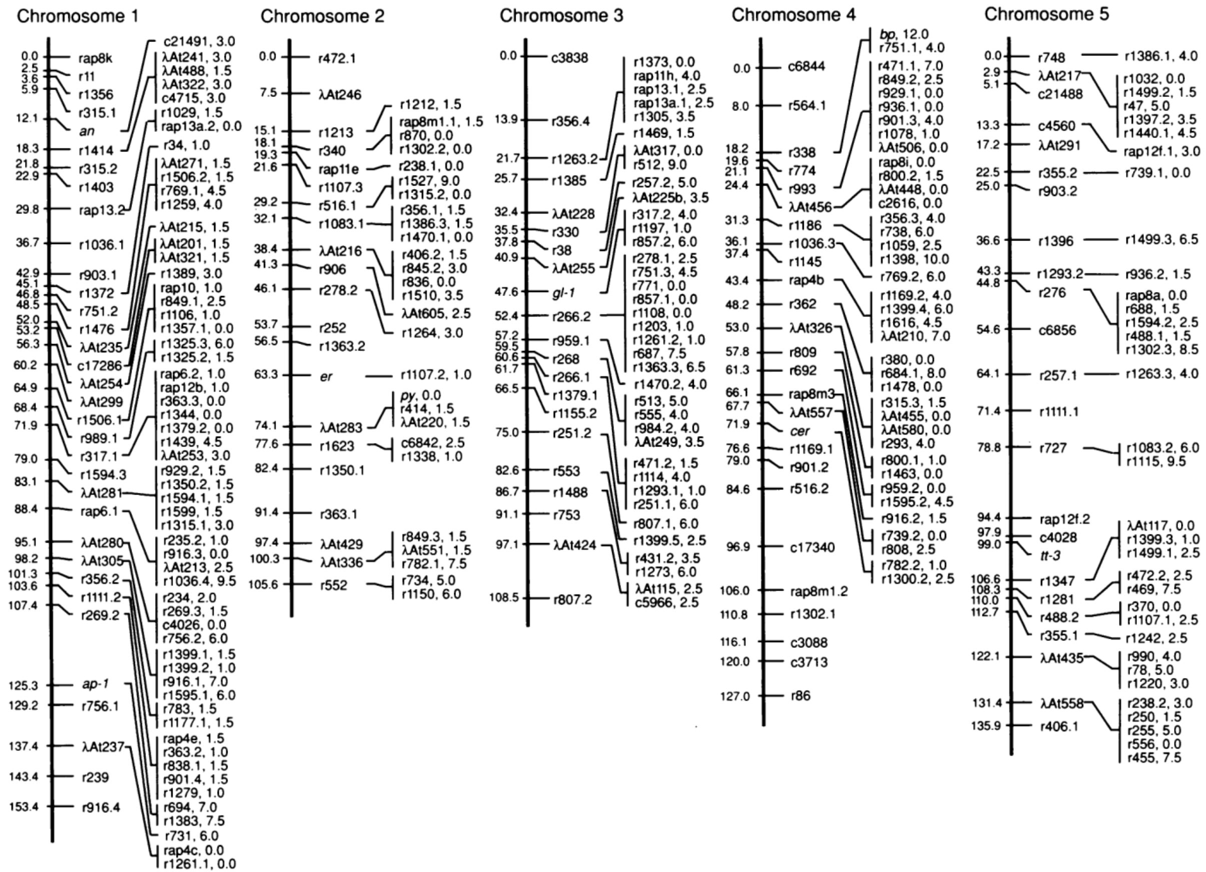


FIG. 3. Linkage map of *A. thaliana*. The 126 markers along the five vertical bars were ordered with LOD score differences (>3.0; see ref. 22). Marker loci listed to the right of each chromosome bar could not be ordered with equal confidence (LOD score differences <3.0). All loci were linked with a LOD score >6.0. The first marker at the top of each chromosome was assigned position 0.0, and approximate marker positions are shown in cM to the left of each chromosome bar. Markers to the right are listed along with an approximate distance in cM from the markers placed on the LOD 3.0 linkage map. Chromosome 1 RAPD markers with numbers >r1300 were placed by local mapping. Phenotypic markers are italicized; RFLP markers are designated either λAt (4) or c (5). The remaining markers are RAPDs.

- **First "dense" RAPD map**
  - Developed by DuPont Compant
    - Developers of RAPD concept and the RAPD map

Mapping Efficiencies of Different Mapping Populations

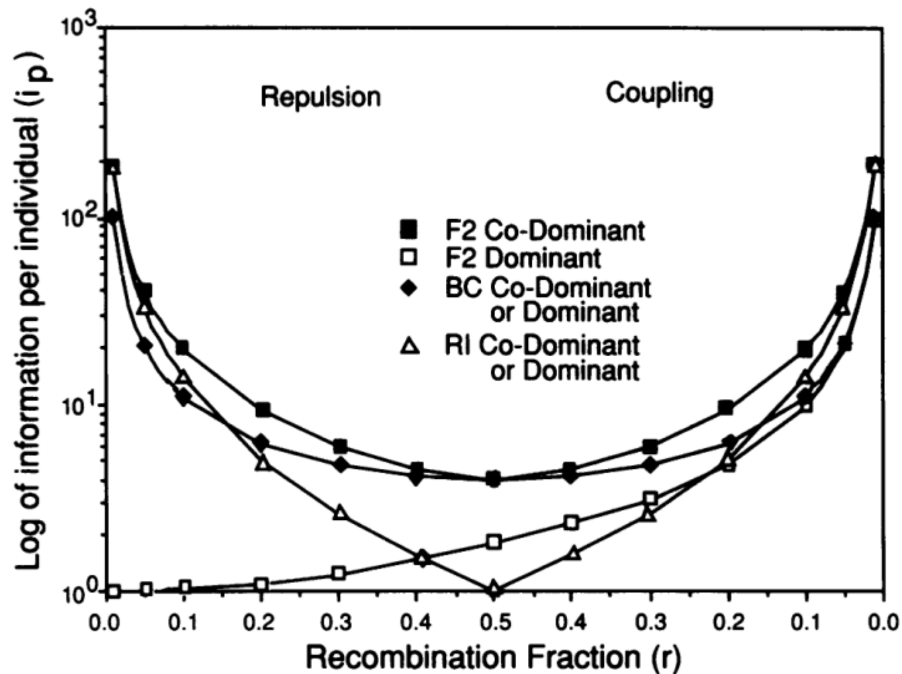


FIG. 5. Theoretical efficiency with which RI, F_2, and backcross (BC) populations can detect recombinants by using either codominant or dominant markers. The amount of information per individual (i_p) in a mapping population is the inverse of the variance divided by population size (28). For an RI population i_p is approximately equal to $2/r(1 + 2r)^2$. Allard (28) previously derived i_p for the other populations shown. The amount of information per individual is represented by the logarithm of i_p and is plotted against the recombination fraction (r) for repulsion- and coupling-phase linkage. An RI population is equally efficient with either codominant or dominant markers and is very efficient for closely linked markers.

Other Mapping Populations

Association Mapping (AM) Population

Limits of traditional bi-parental populations

- Limited number of recombination events
    - F_2 population
        - One round of recombination
    - Recombinant inbred population
        - Two rounds of recombination
- Allele richness of F_2 and RI populations
    - Poor
        - Only alleles of parents sampled
            - Limits discovery of all factors controlling a quantitative trait in a species
- But the advance of F_2/RI populations
    - With limited recombination
        - Fewer markers needed to discover relevant genetic factors

What is an association mapping (AM) population?

- Collection of genotypes from any level of organization from a species
    - Represent the genetic background for which you want to make inferences
        - Wild lines, landraces, market classes, released varieties, breeding lines
        - Population often referred to as a ***diversity panel***
- Arabidopsis
    - Collection of wild samples from throughout the world
        - 25 lines, four populations each
            - PLoS Biology (2005) 3:e196
- Maize
    - 92 inbred lines
        - 12 stiff stalk, 45 non-stiff stalk, 35 tropical + semi-tropical
            - Nature Genetics (2001) 28:286
- Major benefit of AM
    - ***Samples many more recombination events than other populations***
        - Greater resolution than F_2 or RI populations
- Resolution depends upon the linkage disequilibrium in the population
    - But the disadvantage is
        - You need many markers to find meaningful associations
- ***Population size used today for plant AM experiments***
    - 200-300 to 400-500 lines

MAGIC Populations

MAGIC

- **M**ulti-parent **A**dvanced **G**eneration **I**ntercross **P**opulation

Typically

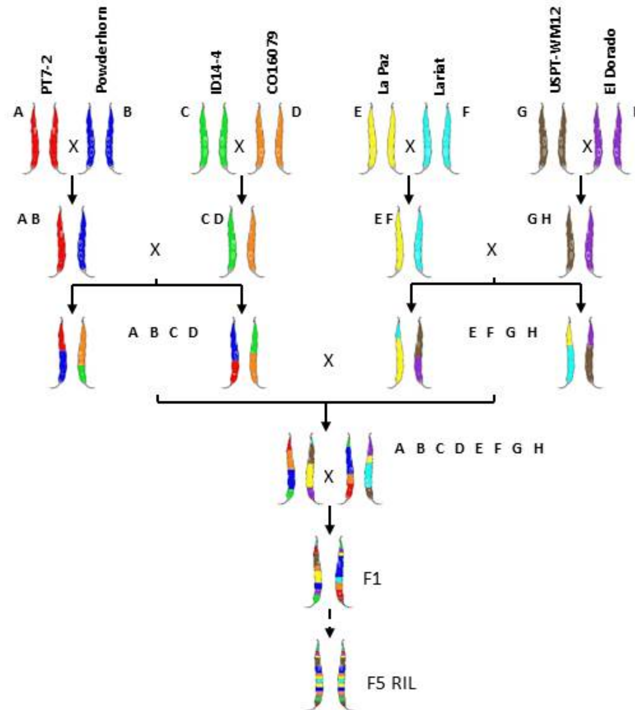
- 8 parents
  - Landrace or established cultivars
- 4 two-crosses at generation 0
- 2 four-way crosses at generation 1
- 1 eight-way crosses at generation 2
- Then multiple rounds of selfing to produce individual lines

Advantages

- Population can be a used for genetic analysis and cultivar development
  - Population can be selected for lines with large number of favorable alleles
- Lines segregating for eight alleles
  - Many different recombination types
  - Minor allele frequency is maximum of 12.5%
- Can be evaluated as an association mapping population using genome-wide association study (GWAS) methods

MAGIC Population Development

Multi-parent Advanced Generation Inter-cross population



Nested Association Mapping (NAM) Population

How is a NAM population created?

- Developed from crossing
  - One common parent to
    - Multiple parents of diverse origin
      - Parents contain relevant diversity specific to the trait(s) of interest to the research community
- It is a community resource
  - High development cost
  - Parents must be carefully chosen
- Maize example
  - B73: common parent
    - Reference genotype used for the genome sequencing project
  - 25 other parents
    - Represent the diversity of the species
  - 200 lines developed per cross
    - 5,000 total lines
      - Large number of lines give better mapping resolution
- Genetics (2009) 183:1525

Advantages

- Like bi-parental population
  - Uses new recombination events created by the cross
    - Fewer markers than GWAS to discover QTL
- Like AM population
  - Uses many recombination events
    - These derived from the many crosses used to develop the population

Diversity or Phylogenetic Populations

Used to determine the phylogenetic relationship among individuals

- Define patterns of relatedness
- Selection of parents for breeding
- Determine ancestral origin of population
- Define a gene tree
    - Origin of a gene in a lineage
- Define a species tree
    - What is the relationships of members of a species
        - Within a species
        - Within a genera
        - Within a family

Considerations

- ***The collection of genotypes should represent the lineage that is of concern to the study***

Bulk Segregant Analysis (BSA)

Efficient method of quickly finding RAPD markers linked to a gene

- Create two bulks
    - Bulk 1: Shared phenotype
        - Disease resistant
    - Bulk 2: Shared phenotype
        - Disease susceptible
- For most RAPD loci (specific primer)
    - Both pools will have the same amplification product
    - Why?
        - They have a combination of the allele that amplifies and the one that does not amplify
- For a rare RAPD loci (specific primer)
    - An amplification product will only be observed in one pool
        - That primer defines a locus linked to the gene controlling the phenotyope.

Major paper describing the procedure (Michelmore et al. 1991. Proc. Natl. Acad. Sci. 88:9828

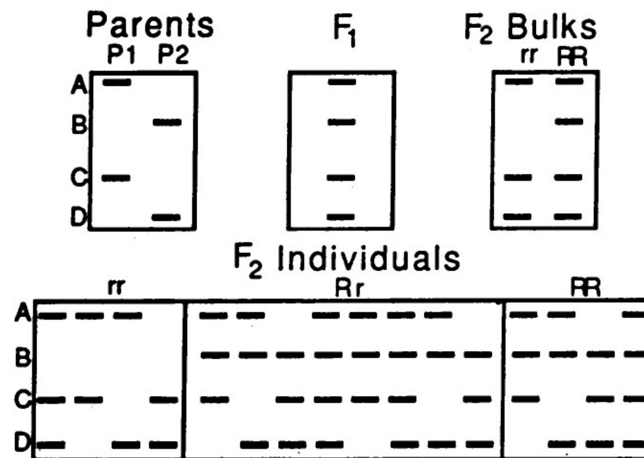


FIG. 1. Genetic basis of bulked segregant analysis. The schematic shows genotypes of four RAPD loci (A–D) detected by a single primer in two parents (P1 and P2), their F_1 and F_2 progeny, and bulks derived from F_2 individuals homozygous for resistance or susceptibility. The dominant allele at locus B is linked in cis to the *R* allele and therefore is polymorphic between the bulks. The other three loci that are polymorphic between the parents are unlinked to the resistance locus and therefore appear monomorphic between the bulks. This is an interpretation of the pattern obtained with primer OPF12 in Fig. 4.

- The dominant allele is linked in cis to the resistant allele
  - The linked band is found in the homozygous and heterozygous resistant plants
- For some linked markers, the presence of the band is linked to the recessive allele
  - In this case only the resistant lines will have the band

Other uses of BAS

- Mapping QTL loci
    - Bulk on the extremes of the phenotypic distribution
    - Screen with dominant (RAPD) marker
- QTL-seq
    - Bulk on phenotypic extremes
        - Qualitative or quantitative traits can be evaluated
    - Sequence the pools
        - Compare sequence variation for linked loci