Road Map Analogy of Molecular Mapping

Let’s find Fargo, your new home for a few years.

Let’s think of the:

- **World** = **Whole Genome**
- **Country** = **Chromosome**
- **State** = **Arm of Chromosome**
- **Border of the State** = **Subsection of Chromosome Arm**
- **City** = **Gene**
- **Streets and Homes in the City** = **Parts of the Gene**

Road Map Analogy of Molecular Mapping

Finding Fargo!!!

1. Locate the United States (= single chromosome; the world is the genome with many chromosomes = countries)

WASHINGTON
OREGON
IDAHO
MONTANA
NORTH DAKOTA
MINNESOTA
WISCONSIN
MICHIGAN
MAINE
VT
NH
NEW YORK
MASS
CONN
RHODE ISLAND
WYOMING
SOUTH DAKOTA
NEVADA
CALIFORNIA
UTAH
COLORADO
NEBRASKA
IOWA
ILLINOIS
INDIANA
OHIO
PENNSYLVANIA
NEW JERSEY
DELAWARE
MARYLAND
DC
WEST VIRGINA
VIRGINIA
KANSAS
MISSOURI
KENTUCKY
NORTH CAROLINA
TENNESSEE
SOUTH CAROLINA
ARIZONA
NEW MEXICO
OKLAHOMA
ARKANSAS
ALABAMA
GEORGIA
MISSISSIPPI
LOUISIANA
TEXAS
FLORIDA
ALASKA
HAWAII

2. Locate North Dakota (Arm of chromosome)

3. Arrive in North Dakota (Subsection of chromosome arm)

4. Arrive in Fargo (In the gene)

- All of the streets and homes would represent the various parts of the gene

Molecular Mapping of Genomes

Molecular Mapping

- Identifies a molecular locus that resides very near or in the gene of interest
    - This locus can be used as a molecular marker for that locus
- The marker lets you know where you are genetically and physically (if the genome of your species has been sequenced)

Molecular Mapping is A Major Goal of Many Genetic Analyzes

- To locate the genetic and physical position of a gene in the genome
    - Enables
        - The eventual cloning of that gene
        - Development of a functional marker that
            - Resides near the gene
            - Cosegregates with the gene
            - But is not part of the gene

What principle of Mendelian genetics is used for mapping?

- Linkage
    - Linkage represents the genetic relationship between any two loci in a genome

Locus

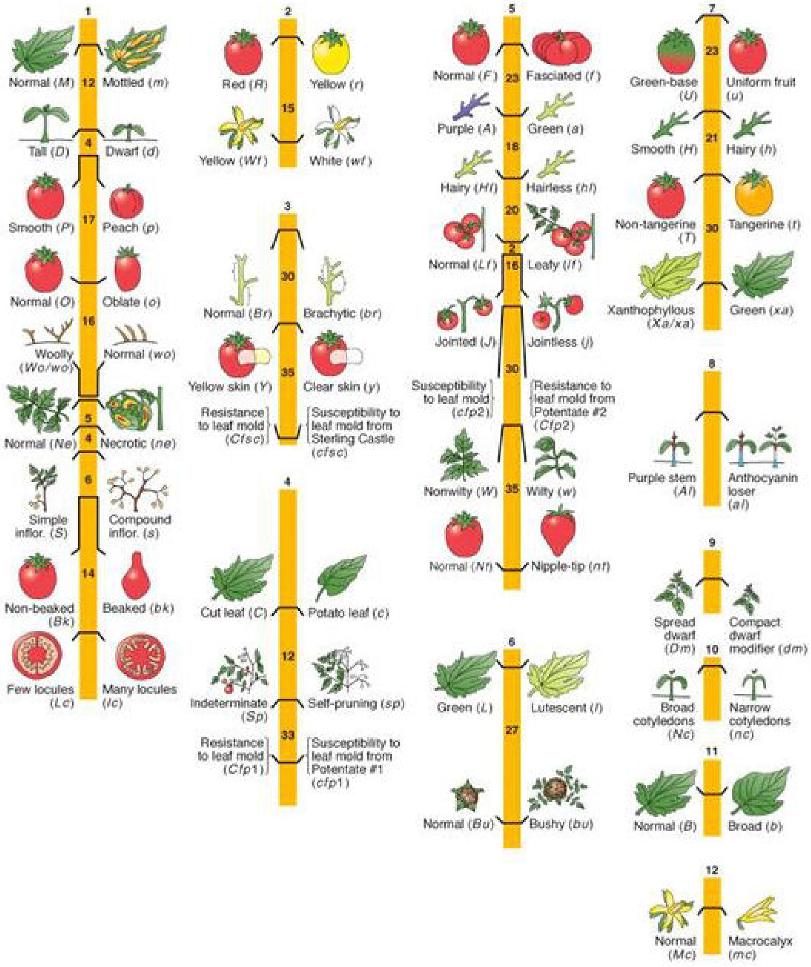
- A genetic position in a genome
    - This could be a
        - Gene
        - Molecular marker

Initial Step in Nearly All Current Molecular Genetic Experiments

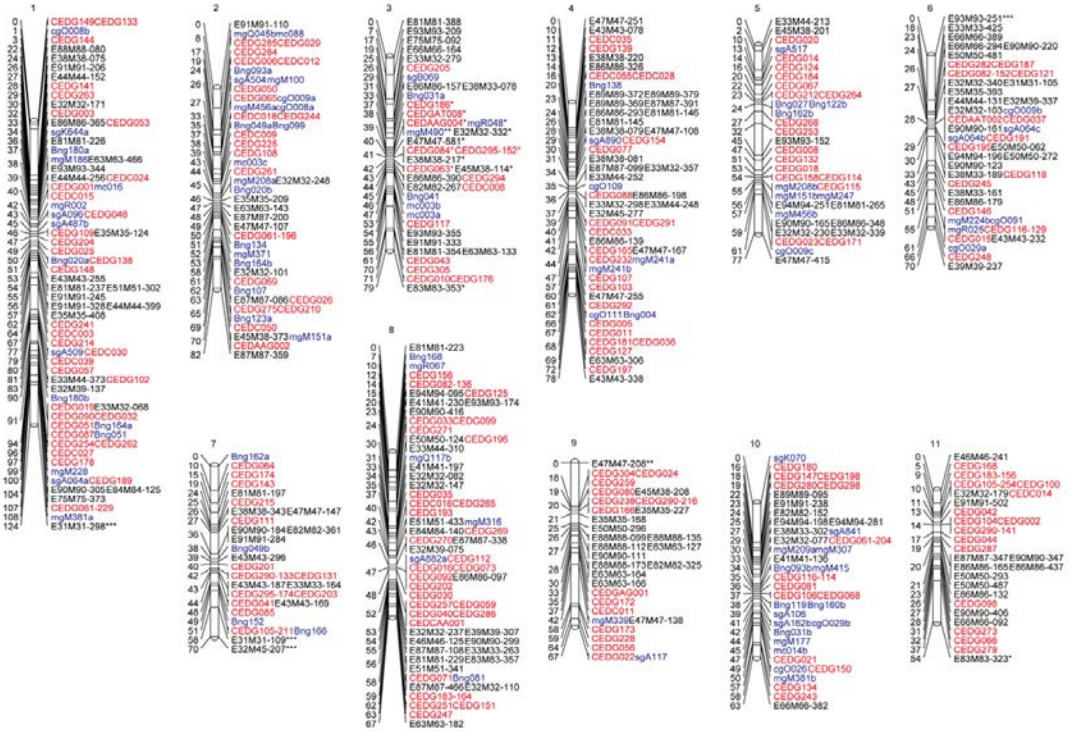
- Develop a linkage map

Phenotypic Maps

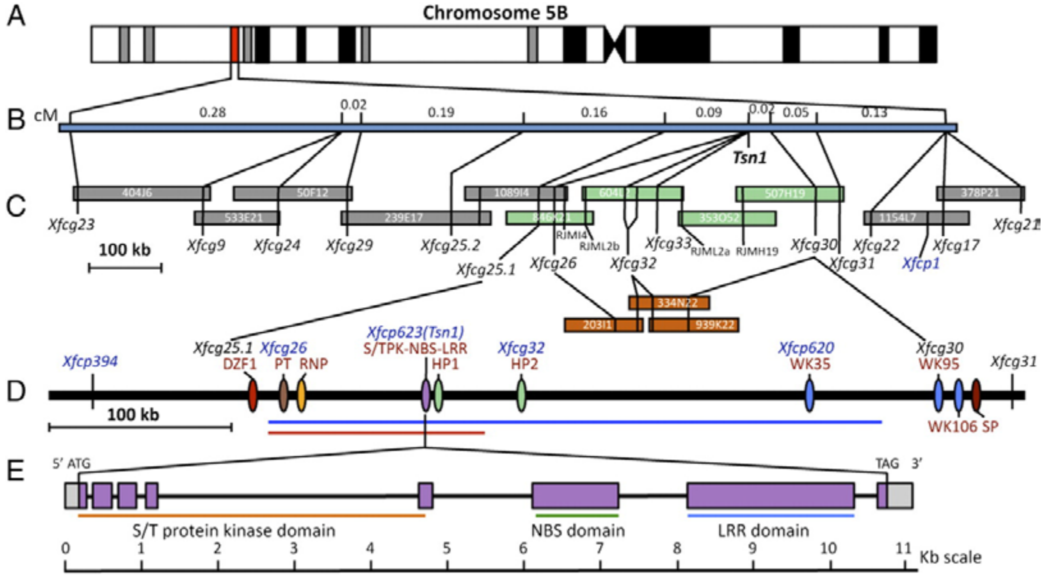
- The first genetic maps that were developed
  - Long term research effort
    - Multiple populations were required
      - Why???
      - Impossible to have all of the loci segregating in one population
      - The plant would be very weak and probably not produce seed



Genetic Map – *Vigna angularis*



Detailed Physical Map of a Genetic Region



What is required to develop the molecular map???

- **Molecular markers!!!!**

Classes of Molecular Markers

Isozymes

- Gel-based approach
- Protein analyzed
- Allelic difference detected
  - Alleles migrate at different rate in starch gel
- Locus-by-locus approach

RFLP – R̲estriction F̲ragment L̲ength P̲olymorphism

- Hybridization-based approach
- RFLP recognized by specific clone/enzyme combinations
- Locus-by-locus approach

RAPD – R̲andomly A̲mplified P̲olymorphic D̲NA

- Gel-based approach
- PCR amplification of fragment
- 10mer oligonucleotides recognize inverted repeats
  - Fragment between the inverted repeats is amplified
- Genome-wide approach

AFLP – A̲mplified F̲ragment L̲ength P̲olymorphism

- Gel- or capillary-based approach
- Combines restriction disgestion with PCR amplification
- Selectively amplifies a subset of possible genomic fragments
- Genome-wide approach

Microsattelites

- Gel- or capillary-based approach
- Mostly based on differences in number of di- or tri nucleotide repeats at a specific locus
- Locus-by-locus approach

SNP – **S**ingle **N**ucleotide **P**olymorphism

- Multiple detection procedures
- Based on nucleotide differences between two alleles
- Locus-by-locus approach

Isozymes

Enzyme Systems

- Rbcs – Ribulose bisphosphate carboxylase, small subunit
- Skdh – Shikimate dehydrogenase
- Prx – Peroxidase
- Me – Malic enzyme
- Mdh – Malate dehyrogenase
- Diap – Diaphorase
- Lap – Leucine aminopeptidase

Matrix

- Starch
- Typically five samples per gel

Scoring Protocol

- Most frequent allele is given the designation “100”
- Other alleles designated by distance in millimeters from the most frequent allele
- Allele nomenclature examples from *P. vulgaris*
    - o Me^{98} – 2 millimeters slower in the gel
    - o Me^{100} – most frequent allele
    - o Me^{105} – 5 millimeters faster in the gel

RFLP Clones, Enzymes, and Informative Hybridizations

1. Source of clones

a. **Random clones** – poor choice because they often contain repetitive DNA

b. **cDNA clones** – contain only expressed sequences; often low copy number

c. **PstI clones** – based on the concept that expressed (low copy number) genes are undermethylated; therefore the methylation sensitive enzyme PstI will only cut in regions where expressed genes reside

2. Enzymes – need to screen parents by digestions with a series of enzymes to find polymorphic hybridizations

3. Informative hybridizations – those specific restriction enzyme/clone combinations that are polymorphic will be informative for mapping a segregating population

Restriction-modification system

Restriction enzyme - cleaves DNA at a specific sequence
Methylase - protects the host DNA from being cleaved

The *E. coli* Type RI (*Eco*RI) restriction enzyme site is:

```
      ↓  m
5' - G A A T T C - 3'
3' - C T T A A G - 5'
           m  ↑


        ↓ (Remember, the enzyme will not cut
        ↓  if the 3' A is methylated.)


5' - G - 3'             5' - A A T T C - 3'
3' - C T T A A - 5'             3' - G - 5'
```

Restriction enzyme = *Eco*RI cuts between the G and A. in the sequence GAATTC
Methylase = *Eco*RI methylase protects the *Eco*RI by adding a methyl group to the 3'-adenine

How the system works
1. Growth of foreign DNA (such as virus DNA) is **restricted** in the cell by the restriction enzyme
2. The bacterial DNA is **modified** by the methylase to prevent cleavage by the restriction enzyme.

Methylation of Plant DNA and its Effect on Restriction Digestion

Grenbaum et al. Methylation of cytosines in higher plants
Nature 297:86 August 1981

5-methyl cytosine (5mC) is found to be a component of plant DNA much more frequently than animal DNA

	% Cytosine as 5mC
Animals	2-7
Plants	>25

Why???

1. 5mC occurs at 70-80% of the 5'-CG-3' dinucleotides. What are the occurrences of the dinucleotide in the two kingdoms?

	% dinucleotides as 5-'CG-3
Animals	0.5-1.0
Plants	3.4

2. 5mC also occurs at the 5'-CXG-3' sequence in plants but does not occur in animals. How often are these sequences methylated in plants?

Sequence	% C methylated
5'-CAG-3	80
5'-CCG-3	50

Do Not Use These Enzymes to Analyze Plant DNA

1. Site has 5'-CXG-3'

*Pst*I	5'-C*TGCAG-3'
*Pvu*II	5'-CAGC*TG-3'
*Msp*I	5'-C*CGG-3
*Eco*RII	5'-CC*WGG-3' (W = A or T)

2. C at or near end of site;if next base in DNA is G, it will be methylated

*Bam*HI	5'-GGATC*C-3
*Kpn*I	5'-GGTACC*-3

Use These Enzymes Instead

*Dra*I	5'-TTTAAA-3'
*Eco*RV	5-'GA*TATC-3
*Eco*RI	5-GAA*TTC*-3'
*Hind*III	5'-A*AGC*TT-3'
*Xba*I	5'-TC*TAGA*-3'

RAPD Markers

- A PCR-based marker system

- Amplifies inverted repeats in the genome

- Can generate larger number of markers than RFLP in a short period of time

- Popular because easy to do and does not require radioisotopes

AFLP Procedure

1. Digest sample DNA with restriction enzymes *Eco*RI and *Mse*I.

```
5'----GAATTCN------------------------NTTAA----3'
3'----CTTAAGN------------------------NAATT----5"


       AATTCN------------------------NT
           GN------------------------NAAT
```

2. Anneal *Eco*RI and *Mse*I adaptors to restriction products.

```
??????AATTCN-------------------------NTTA??????
??????TTAAGN-------------------------NAAT??????
```

(**??????** = unknown sequences that are unique for the companies primers)

3. Preselect products by PCR amplification with "*Eco*RI + A" and "*Mse*I +C" oligonucleotide primers.

```
EcoRI Primer+A
????????AATTCN---------------------NTTA????????
????????TTAAGN---------------------NAAT????????
                                   C+MseI Primer
```

4. Selectively amplify preselect PCR products by using "*Eco*RI + 3" and "*Mse*I +3" oligonucleotide primers.

```
EcoRI Primer+AAC
????????AATTCA----------------------GTTA????????
????????TTAAGT----------------------CAAT????????
                                 AAC+MseI Primer
```

5. Separate fragments by denaturing polyacrylamide gel electrophoresis

Microsatellites

Also called:

- SSR = Simple Sequence Repeats
- SSLP = Simple Sequence Length Polymorphisms
- STMS = Sequence Tagged Microsatellites)

What are they?

Typically repeated di-, tri or tetranucleotide sequences

Examples:

- AG_4 (AG AG AG AG)
- CGA_3 (CGA CGA CGA)
- $GATA_5$ (GATA GATA GATA GATA)

How are they discovered?

- Sequenced genes in databases searched for repeats
- Repeat olignucleotide probes used to screen random clone library

Rice example [Mol Gen Genet (1996) 252:597]:

Library:

- 300-600 bp fragments generated by shearing DNA
- fragments cloned into plasmid vector

Probes:

- GA_{13}
- CGG_9
- ATC_9
- ATT_9

The Microsatellite Application

- Primers highly specific to sequences flanking the repeat are designed
- Individual DNA samples are amplified
- Products compared by polyacrylamide or agarose gels using staining producedures
- Recently fluorescently labeled primers are used and products are analyzed with laser technology (gel or capillary)
- Differences in two samples are represented by size differences of amplified fragments
- The size difference is the polymorphism
- Single-base pair polymorphisms can be detected

Advantages of Microsatellites

Genetic

- Many loci in the genome (goal: 30,000 in humans)
- Randomly distributed in a genome
- Extensive polymorphism within a species
- Many act as codominant markers

Technical

- Generally reproducible from lab to lab
- Small amounts of target DNA needed
- Can be automated
- Mulitplexing possible

Disdvantages of Microsatellites

Genetic

- Null alleles at a specific locus result in a dominant marker and heterozygotes can not be scored

Technical

- Very high development costs

Arabidopsis Microsatellite Summary

Bioinformatics (2004) 20:1081-1086

- 93% of microsattelites are trinucleotides
- Repeats less frequent in coding than non-coding regioins
- Microsatellite more abundant in 5' region of gene
- AG and AAG were the most frequent repeats
- Studied 1140 full length genes with an AG or AAG repeat in 5" region
  - These repeats often associated with function of gene
  - Based on proximity to beginning of first exon
    - These genes more often performed a molecular activitythan genes without a repeat

Modern Approach to Microsatellite Discovery

Collect sequence data

- EST
- Will be in genes that are in the low copy region of the genome

Random genome reads

- May be low copy or high copy regions of the genome

Soybean EST Discovery

- Song et al. Crop Science 50:1950 (2010)

Screened

- Whole genome sequence
- ESTs
  - Minimum of 5 copies of the repeat

	Genome sequence	EST sequences
Total	210,990	33,327
Dinucleotide	168,625	20,161
Trinucleotide	38,411	12,440
Tetranucleotide	3,954	636

Most abundant in genome sequence

- (AT)n, (ATT)n, (AAAT)n
  - 61,458

Testing the approach

- 1034 primer pairs developed
- Screened 7 diverse genotypes
  - 94.6% (978) amplified a single PCR product
    - 77.2% (798) were polymorphic

Single Nucleotide Polymorphisms (SNPs)

Single nucleotides are the smallest unit of mutation

Single nucleotides therefore are the smallest polymorphic unit

SNP = single nucleotide polymorphisms

Uses:

- Detect level of variation within a species
- Follow patterns of evolution
- Mark genes
- Distinguish alleles of “disease” genes
- Create designer pharmaceuticals

The SNP Consortium, Ltd

Goal

- Identify 300,000 human SNPs by Dec. 31, 2001

Progress

- 7,365 (Dec. 31, 1999)
- 1.2 million detected (Jan 26, 2002) and mapped in one population
- 1.1 high quality mapped in three populations; 1 SNP every 5 kilobases

Detecting DNA Polymorphisms

DNA molecules greater than 10 base pairs contains essentially the same mass-to-charge ratio

Procedure that separates the molecules based on mass alone will uncover DNA polymorphisms

Current Procedure

- Gel electrophoresis

Emerging Procedures

- Capillary array electrophoresis
- Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)
    - New technology in 2002 that has not really caught on

Advantage of New Procedures

- SPEEEEEEEED

Gel Electrophoresis

Most widely adapted technique for detecting polymorphism.

Principle

- Samples loaded into a gel and allowed to migrate in an electric field
- DNA is positively charged and samples migrate toward the negative pole
- Separation of the molecules is strictly based on size
- smallest fragments move farther in the gel
- They can navigate through the small pores in the gel faster than large molecules.

Types of Gel Matrices

- Agarose
- Polyacrylamide

Separation is function of the polymer concentration

Agarose		Polyacrylamide	
%	**Resolution (kb)**	**%**	**Resolution (bp)**
0.9	0.5 - 7.0	3.5	1000 - 2000
1.2	0.4 - 6.0	5.0	80 - 500
1.5	0.2 - 3.0	8.0	60 - 400
2.0	0.1 - 2.0	12.0	40 - 200

What polymer do you choose?

Polymer choice is based on the size range of fragment

Agarose

- RFLP
- RAPD

Polyacrylamide

- Microsatellites
- AFLPs

Observing the Polymorphisms

Dye Detection

- Ethidium bromide (Agarose)
  - RAPD
- Silver nitrate (Polyacrylamide)
  - Microsatellites
  - AFLPs

Authoradiography Detection

- RFLP (more later)
- Microsatellites
- AFLP

Laser Technology Detection

Applied to the microsatellites

Procedure

- PCR primer is labelled with a fluorescent dye
- Samples are separated in a polyacrylamide gel
- Fragments are detected by a laser as they flow through the bottom of the gel
- Computer programs output data as size difference

RFLPs and Southern Hybridizations

- Restriction digested DNA separated in agarose gel
- DNA transferred to a nylon membrane
- Membrane contains a faithful representation of the fragment distribution found in the gel
- Probe is hybridized to the membrane
- Unbound probe is removed by a series of stringent washes
- Membrane is exposed autoradiographic film
- Polymorphisms observed

Capillary Array Electrophoresis

Small DNA fragments are rapidly separated in narrow capillaries.
Heat is lost rapidly from thin capillaries
Thus, molecules can be separated rapidly using high voltage.
High voltage separation greatly speeds sample processing

Capillaries arrays have reached the market

- Eight capillary array (Beckman-Coulter)
- 96 capillary array (Applied Biosystems)

Advantage

- Multipe samples simultaneously introduce into each arrays
- Size detected by laser technology

Advantages and Disadvantages of Different Marker Systems

Marker	Advantages	Disadvantages
Classical	Easily scored; new equipment not needed; traits of economic importance	Time to score; environmental influences; limited for most species; mostly dominant
Isozyme	Easy to apply; codominant; multiple alleles	Limited number of assays available; scored at only one developmental stage
RFLP	Codominant; highly repeatable; maps available for major species	Limited polymorphisms, time consuming; relatively expensive; requires radioactivity
RAPD	Easy technique; rapidly screening; no radioactivity	Low lab-to-lab repeatability; mostly dominant
AFLP	Large number of polymorphisms per sample; easy to apply (once mastered)	Difficult to learn; radioactivity might be used; expensive; time-consuming; mostly dominant
Microsatellites	Highly polymorphic; easy to apply; large number of loci available; mostly co-dominant	Slow and expensive development costs
SNP	High accuracy; potentially a very large number; co-dominant	Expensive to discover; assays are currently expensive

What to consider when choosing a marker system

- Genetic nature of the system
- Radioactivity
- Cost
- Reproducibility
- Time (Discovery and application)
- Efficiency (Time invested vs. breadth of results)
- Is a system in place??? (Map available; technique defined)
- Reason for experiments (Marker development vs. application)
- Amount of polymorphisms
- Technical skill needed

Types of Nucleic Acid Hybridizations

1. **Southern hybridization** - hybridization of a probe to filter bound ***DNA***; the DNA is typically transferred to the filter from a gel

2. **Northern hybridization** - hybridization of a probe to filter bound ***RNA***; the RNA is typically transferred to the filter from a gel

Probe - a single-stranded nucleic acid that has been radiolabelled and is used to identify a complimentary nucleic acid sequence that is membrane bound

The following steps describe the ***Southern transfer procedure***.

1. Digest DNA with the restriction enzyme of choice.
2. Load the digestion onto a agarose gel and apply an electrical current. DNA is negatively charged so it migrates toward the "+" pole. The distance a specific fragment migrates is inversely proportional to the fragment size.
3. Stain the gel with EtBr, a fluorescent dye which intercalates into the DNA molecule. The DNA can be visualized with a UV light source to assess the completeness of the digestion.
4. Denature the double-stranded fragments by soaking the gel in alkali (>0.4 M NaOH)
5. Transfer the DNA to a filter membrane (nylon or nitrocellulose) by capillary action.

Steps in Southern Hybridization Procedure

1. Prepare a probe by nick translation or random, oligo-primed labelling.

2. Add the probe to a filter (nylon or nitrocellulose) to which single-stranded nucleic acids are bound. (The filter is protected with a prehybridization solution which contains molecules which fill in the spots on the filter where the nucleic acid has not bound.

3. Hybridize the single-stranded probe to the filter-bound nucleic acid for 24 hr. The probe will bind to complementary sequences.

4. Wash the filter to remove non-specifically bound probe.

5. Expose the filter and determine:
    a. Did binding occur?
    b. If so, what is the size of hybridizing fragment?

Hybridization Stringency

- Temperature and salt concentrations of hybridization conditions directly affect hybridization results
- The degree of homology required for binding to occur can be controlled by these factors
- Results are directly related to the “degrees below the T_m” at which the hybridizations and washes are performed
- T_m is the melting temperature of the DNA

T_m = 69.3^{o}C + 0.41(% G + C)^{o}C

From this formula you can see that the GC content has a direct effect on T_m. The following examples, demonstrate the point.

T_m = 69.3^{o}C + 0.41(45)^{o}C = 87.5^{o}C (for wheat germ)

T_m = 69.3^{o}C + 0.41(40)^{o}C = 85.7^{o}C

T_m = 69.3^{o}C + 0.41(60)^{o}C = 93.9^{o}C

Hybridizations though are always performed with salt. This requires another formula which considers this fact. This formula is for the ***Effective*** **T_m (Eff T_m)**.

Eff T_m = 81.5 + 16.6(log M [Na^+]) + 0.41(%G+C) - 0.72(% formamide)

Na^+ ion concentration of different strengths of SSC

SSC Content	$[Na^+]$ M
20X	3.3000
10X	1.6500
5X	0.8250
2X	0.3300
1X	0.1650
0.1X	0.0165

Another relevant relationship is a that ***1% mismatch of two DNAs lowers the*** **T_m** ***1.4^{o}C***. So in a hybridization with wheat germ that is performed at T_m - 20^{o}C (=67.5^{o}C), the two DNAs must be 85.7% homologous for the hybridization to occur.

100% - (20^{o}C/1.4^{o}C) = 85.7% homology

Let's now look at an actual experiment.

Wheat DNA
Hybridization at 5X SSC at 65°C
Non-stringent wash: 2X SSC at 65°C
Stringent wash: 0.1X SSC at 65°C

The first step is to derive the Eff T_m.

Eff T_m = 81.5 + 16.6(log 0.825) + 18.5 = 98.6

Next figure out hybridization homology

100 - [(98.6-65.0)/1.4] = 100 - (23.6/1.4) = 83.1%.

Next figure out Eff T_m and hybridization homology for non-stringent wash

Eff T_m = 81.5 + 16.6[log(0.33)] + 0.41(45%) = 92.0°C

% Homology = 100 - [(92-65)/1.4] = 80.7%

Next figure out Eff T_m and hybridization homology for stringent wash

Eff T_m = 81.5 + 16.6[log(0.0165)] + 0.41(45%) = 70.4°C

% Homology = 100 -[(70.4-65)/1.4] = 96.1%

Stringency - a term used in hybridization experiments to denote the degree of homology between the probe and the filter bound nucleic acid; the higher the stringency, the higher percent homology between the probe and filter bound nucleic acid

Mapping and Mapping Populations

Types of mapping populations

a. F2

b. Backcross

c. Recombinant Inbred Lines (RILs; F2-derived lines)

Homozygosity of Recombinant Inbred Lines

RI population	% within-line homozygosity at each locus
$F_{2 \cdot 3}$	75.0
$F_{2 \cdot 4}$	87.5
$F_{2 \cdot 5}$	92.25
$F_{2 \cdot 6}$	96.875
$F_{2 \cdot 7}$	98.4375
$F_{2 \cdot 8}$	99.21875

Value of Recombinant Inbred Population

1. Eternal source materials

2. Phenotypic data collection can be replicated to ensure accuracy

3. Large field trials can be performed to collective quantitative trait data

4. Problem: dominance and epistasis can not be measured because no heterozygotes

Segregation Ratio of Mapping Populations

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

*To score a dominant maker in a backcross population, you must cross the recessive parent with the F_1 plant. Therefore to score RAPD loci 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 RAPD loci.

Other Mapping Populations

Association Mapping (AM) Population

Limits of traditional bi-parental populations

- Limited number of recombination events
    - F2: one round of recombination
    - RI populations: maximum of two rounds of recombination
- Allele richness
    - Poor
        - Only alleles of parents sampled
- Limits discover of all factors controlling a quantitative trait in a species
- But the advantage is:
    - With limited recombination
        - Fewer markers needed to discover relevant genetic factors

What is an association mapping population?

- Collection of genotypes from a species
    - Represent the genetic background for which you want to make inferences
- Arabidopsis
    - Collection of wild samples from throughout the world
        - 25 regions, four populations each
            - PLoS Biology (2005) 3:e196
- Maize
    - 92 inbred
        - 12 stiff stalk, 45 non-stiff stalk, 35 tropical, semitropical
            - Nature Genetics (2001) 28:286
- Major benefit
    - Samples many more recombination events
        - Great resolution
            - Resolution depends upon linkage disequilibrium in sample
- But the disadvantage
    - Need many markers to find meaningful associations
- Size for population used today

o Several hundreds (200-300 individuals)

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
- Often a single populations is developed per species
    - High resource cost, so
        - Choice of parents is important
- Maize example
    - B73: common parent
        - Sequenced by maize genome project
    - 25 other parents
        - Represent diversity in maize
            - 200 lines per cross (5,000 total)
    - Science (2009) 325:714
- Large number of populations give better resolution
    - Genetics (2009) 183:1525

Advantages

- Like bi-parental population
    - Uses current recombination events created by the cross
    - Fewer markers than AM will discover the QTL
- Like AM population
    - Uses many recombination events
    - Based on the many crosses used to make the population members

Diversity or Phylogenetic Populations

Used to determine the relationship among individuals

- Define patterns of relatedness
- Selection of parents for breeding
- Determine ancestral origin
- Define 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

- Should represent the lineage that is of concern to the study

Specialized Mapping Topics

Bulk Segregant Analysis

Useful for targeting a specific genomic region

Create two DNA bulks

- one contains homozygous dominant individuals
- other contains homozygous recessive individuals

Perform a molecular marker analysis

The bulks are equally random for all regions of the genome except that which contains your gene of interest

Any difference should be linked to the gene controlling the trait you bulked upon

Sequence Tagged Sites (STS)

Defined by a pair of PCR sites obtained from DNA sequencing

Each site is usually 18-20 nucleotides long

Amplification is more specific because of the size of the primer that anneals to one end of the STS

May require subsequent restriction digestion to define a polymorphism

Reduces lab to lab variability seen with RAPDs

Arabidopsis STS Primers (called CAPS)

- two per chromosome
- allows a rapid mapping of new mutant to a specific chromosome