Mutations, Meioses, and Maps

- Adaptive success of a population requires genetic variation
- Re-assortment of traits increases available variation
- Meiosis is the cellular mechanism of re-assortment
- Genetic linkage is a consequence of meiotic recombination
- Extrapolation of basic principles allows analysis of complex traits by linkage, linkage disequilibrium and association studies
- Genome sequence expedites mapping and association studies

I. Mutation

- Kinds of events
- Mutation rates
- Classes of allele
- Screens for induced mutations

Mutagenic Events

- Point mutations (most frequent class in Hs; avg. heterozygosity ~1/kb)
- Chromosomal rearrangements

  (~1/250-500 live births in Hs; account for ~50% of spontaneous abortions)

  Somatic translocations in cancer: Bcr-Abl leukemia
- Mobile elements (transposons, retroviruses, others)
- Repeat expansions

Classes of Allele

H.J. Muller, 1930's

- **amorph**: complete loss of function
- **hypomorph**: reduction of function
- **hypermorph**: increase of normal function
- **neomorph**: gain of novel function
- **antimorph**: antagonism of normal function (dominant negative)

Chromosomal Rearrangements

- Translocation
- Inversion
- Deletions and Duplications
  - intragenic (CMT1A)
  - contiguous gene syndromes (DiGeorge, BWS)
  - aneuploidy and segmental aneuploidy (Down's)

Mobile elements

- Transposons are major mutagens in many organsims (and can drive speciation)
- Endogenous retroviruses account for up to 20% of spontaneous mutations in mouse, but are far less common in man*

*based on both observed mutations and nucleotide substitution analysis of genome sequence

Repeat Expansions Show Anticipation

- Expansion target may be coding (HD, DRPLA, SCA, etc.), regulatory (FA), or fragile sites (FMR)
- Repeat length has a minimum threshold for disease
- Probability of expansion is length-dependent component

Repeat Expaansions Show Anticipation

- Expansion target may be coding (HD, DRPLA, SCA, etc.), regulatory (FA), or fragile sites (FMR)
- Repeat length has a minimum threshold for disease
- Probability of expansion is length-dependent component

Repeat Expaaaansions Show Anticipation

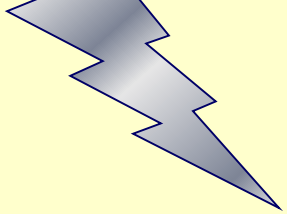
- Expansion target may be coding (HD, DRPLA, SCA, etc.), regulatory (FA), or fragile sites (FMR)
- Repeat length has a minimum threshold for disease
- Probability of expansion is length-dependent component

Repeat Expaaaaaaaansions Show Anticipation

- Expansion target may be coding (HD, DRPLA, SCA, etc.), regulatory (FA), or fragile sites (FMR)
- Repeat length has a minimum threshold for disease
- Probability of expansion is length-dependent component; increasing penetrance or severity in next generation

Repeat Expaaaaaaaansions Show Anticipation

- Expanded polyglutamine creates a toxic misffolded protein (neomorphic) that has low turnover; dominant disorders (HD, DRPLA, SCA)
- Expanded GAA in intron of Fredreich's ataxia gene leads to aberrant processing, recessive disorder



Mutation happens

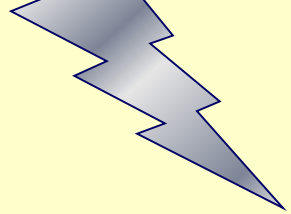
Spontaneous mutation rates μ, per locus per gamete:

E. coli in rich broth: 10^{-7} to 10^{-9}

Humans, Mice, Other metazoa: 10^{-4} to 10^{-6}

Example: Neurofibromatosis type 1A (NF1)
(AKA: von Recklinghausen disease)
Autosomal dominant inheritance
1/3000 live births
1/6000 transmitted gametes
~1/2 are de novo mutations

⇨ μ ~ 1/12000 gametes has new NF1 allele



Mutation happens

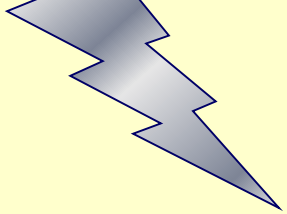
Spontaneous mutation rates μ, per locus per gamete:

E. coli in rich broth: 10^{-7} to 10^{-9}

Humans, Mice, Other metazoa: 10^{-4} to 10^{-6}

Simple sequence repeats: ~10^{-3} to 10^{-6} (molecular)

Example: Huntington's Disease
- Autosomal dominant
- Expansion of $(CAG)_n$ encoding polyglutamine to beyond ~40 residues creates toxicity (neomorph)
- 1 x 10^{-6}
- Anticipation (probability, onset, severity)



Mutation happens

Spontaneous mutation rates μ, per locus per gamete:

E. coli in rich broth: 10^{-7} to 10^{-9}

Humans, Mice, Other metazoa: 10^{-4} to 10^{-6}

Simple sequence repeats: ~10^{-3} to 10^{-6}

Current World Population: 6.1 x 10^{9}
(US Census Bureau 2/23/01)

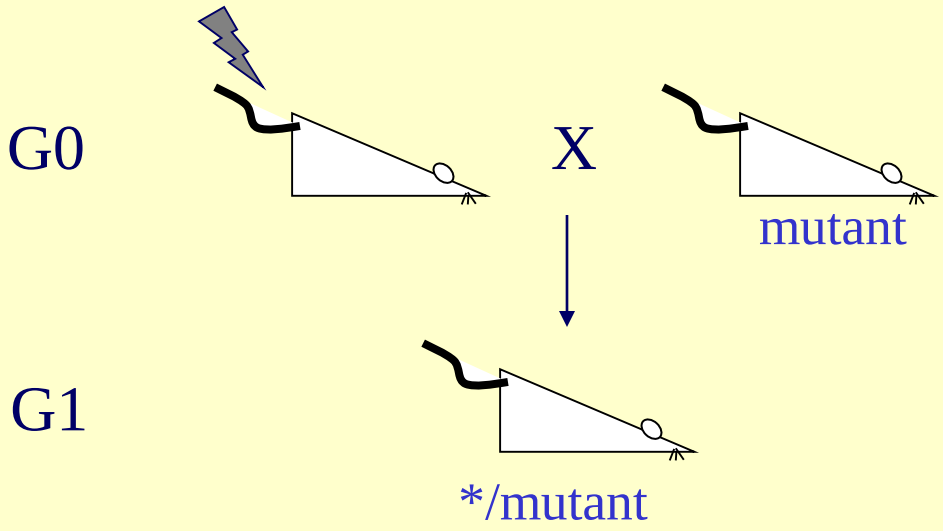
Human reverse genetics: seek and ye shall find

Chemical Mutagenesis

Man's inhumanity to DNA

- Different agents favor different kinds of lesion: x-ray, diepoxides, alkylating agents
- Toxicity vs. Mutagenicity
- Ethylnitrosourea in mice: μ up to 1-3 x 10^{-3} per locus per gamete
- Large-scale screens in progress (NIH, others)
- Saturation screens in yeast, flies and worms

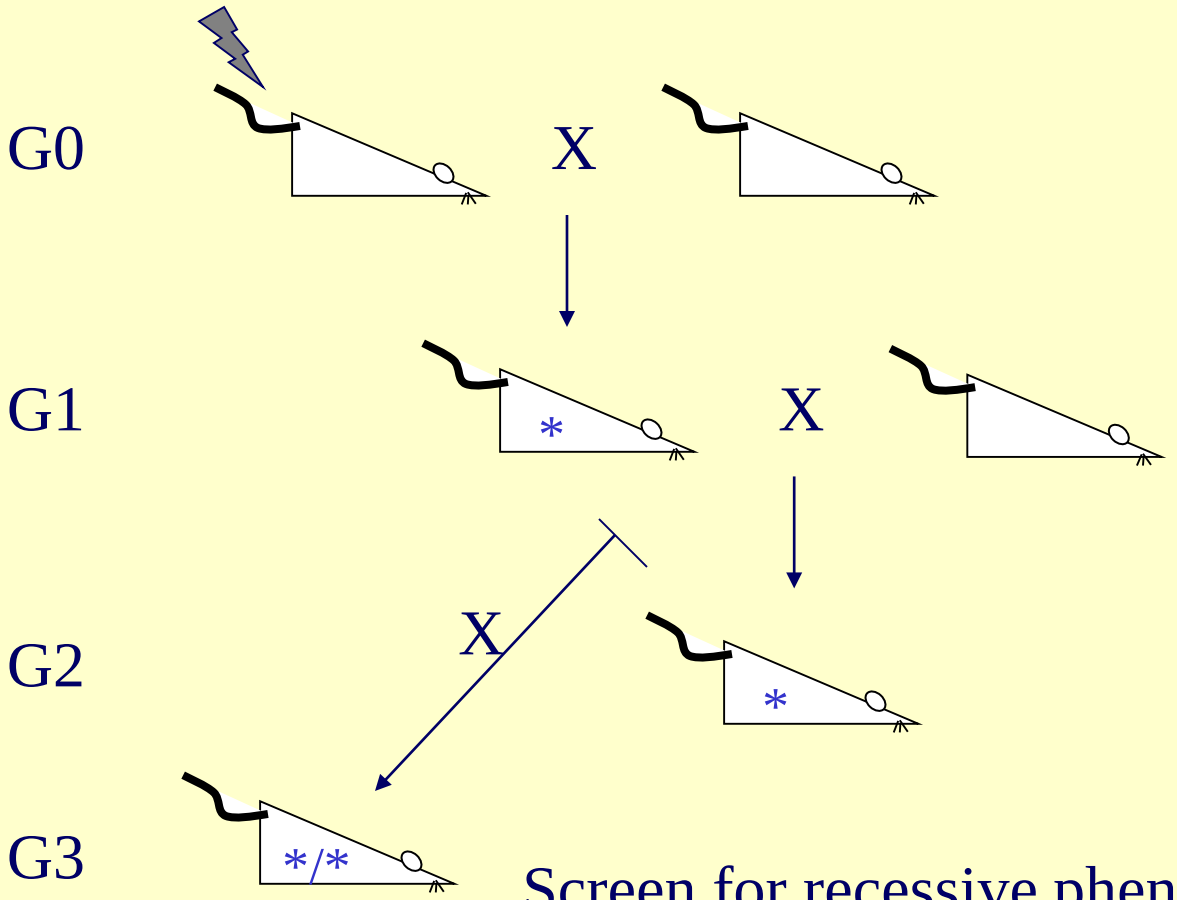
G1 Screen



*Screen for **Genotypes***

- *heteroduplex*
- *sequence*

G3 Screen



Screen for recessive phenotype

II. Meiosis & Mapping

- Simple linkage
- Positional cloning
- Recombination hotspots
- Complex traits and QTLs
- Linkage disequilibrium

(Yes, even if the genome is sequenced)

Meiosis produces 1 crossover per chromosome arm (average)

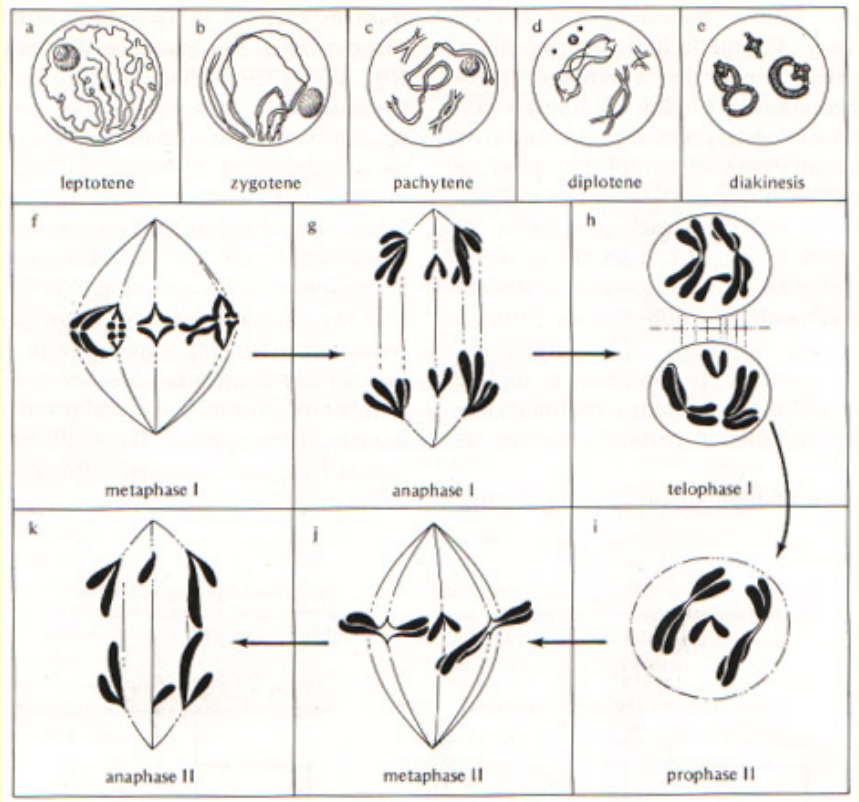
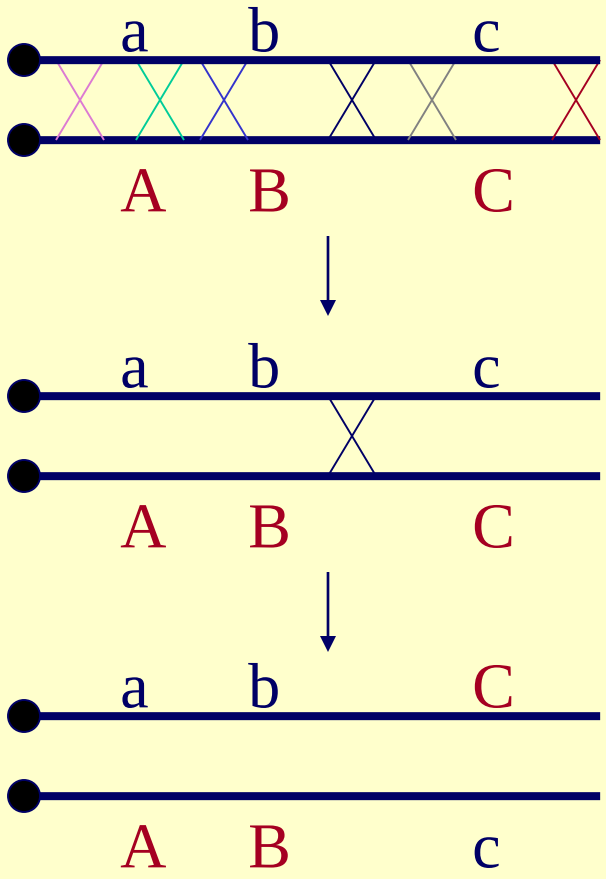


Figure 2–5

Principal stages of meiosis. (From Wolfe, after Lewis and John.)

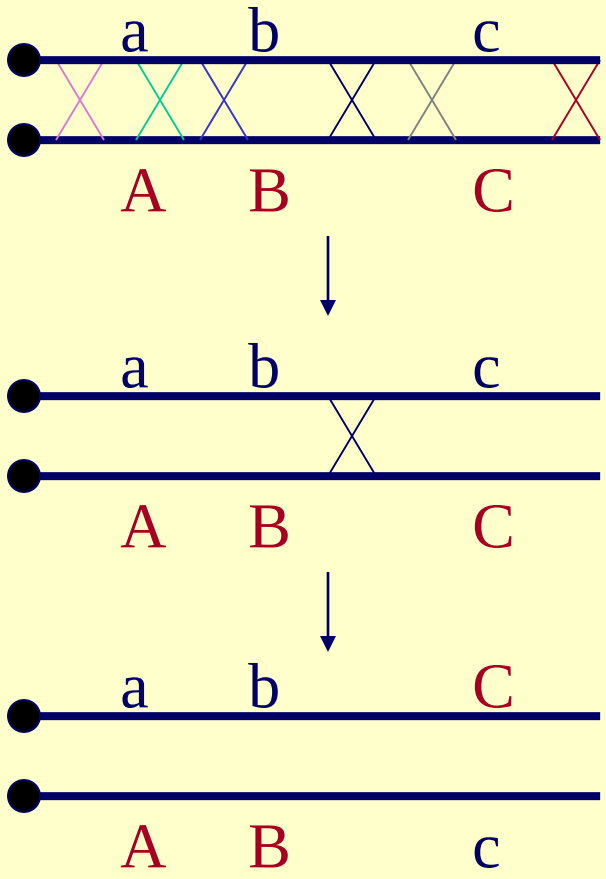
Organism	Mb	kb/cM	cM/arm
Human	3000	1000	78
Mouse	3000	2000	80
Fly	180	500	54
Worm	100	300	25
Yeast	13.5	6	75

Consequences of Meiotic Exchange



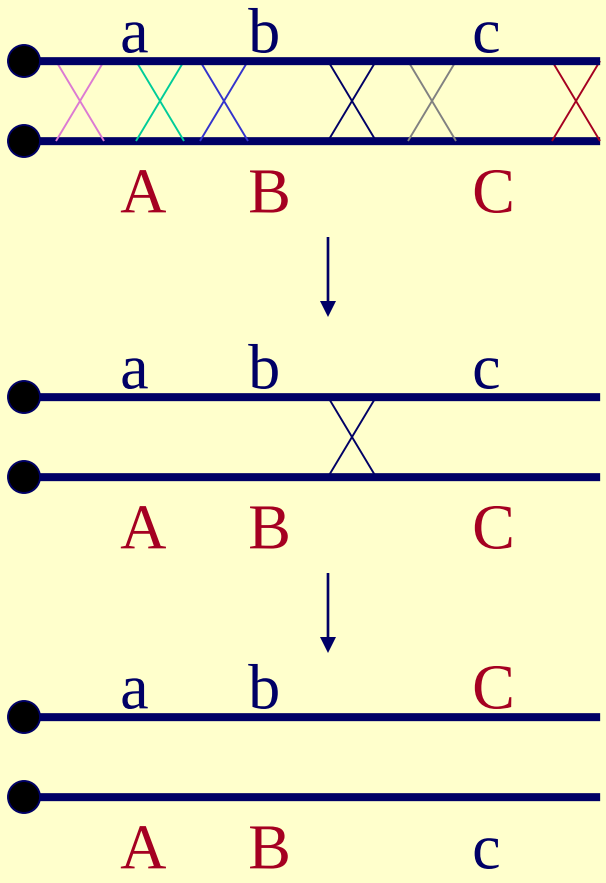
- Crossovers require chromosome pairing
- Crossovers initiated at many sites
- *Interference* prevents closely spaced crossovers
- Average 1 crossover per chromosome arm per gamete

Consequences of Meiotic Exchange



- Map *distance* inferred from number of interval recombinants
- Map *order* inferred by minimizing DCOs
- Information content is highest at a marker and decreases with distance

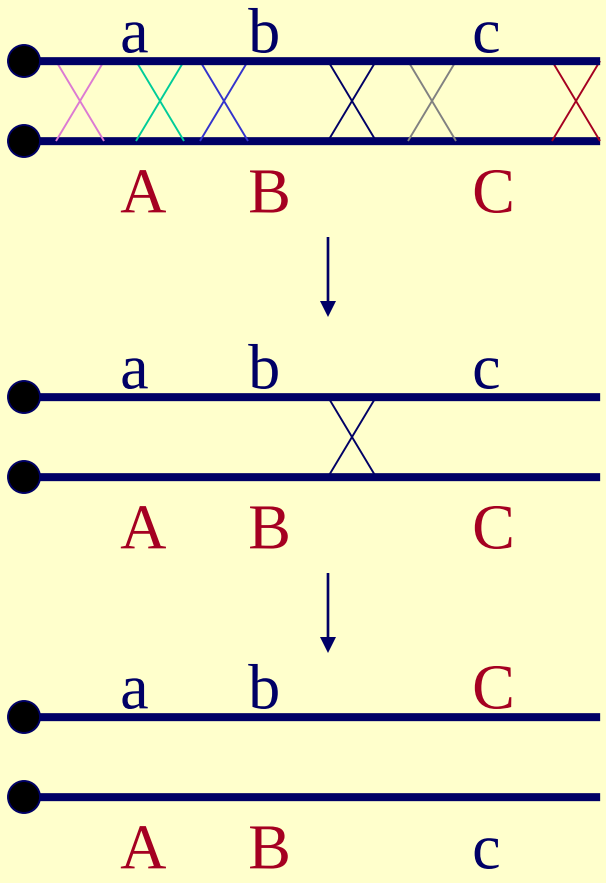
The 3 Factor Cross



Test cross to abc/abc:

Offspring	Number
ABC	354
aBC	41
AbC	8
ABc	88
Abc	35
aBc	10
abC	96
abc	368

The 3 Factor Cross

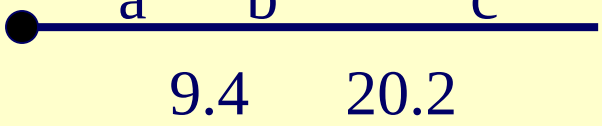


Test cross to abc/abc:

Offspring	Number	
ABC	354	parental
abc	368	
Abc	35	SCO 1
aBC	41	
ABc	88	SCO 2
abC	96	
AbC	8	DCO
aBc	10	

The 3 Factor Cross

Map (in cM):



first genetic map:
A. H. Sturtevant, 1913
(Drosophila X chromosome)

(undergraduate thesis)

Test cross to abc/abc:

Offspring	Number	
ABC	354	parental
abc	368	
Abc	35	SCO 1
aBC	41	
ABc	88	SCO 2
abC	96	
AbC	8	DCO
aBc	10	

Genome-Wide Mapping

- Allows mapping of multiple genes affecting a single trait
  - *quantitative* traits loci (each gene contributes to extent of trait)
  - *qualitative* traits with multigenic threshold (each gene contributes to probability of trait)
  - *epistatic* effects (effect of one gene is dependent on another)

Genome-Wide Mapping

- Power depends on density and information content of markers
- Likelihood of map order or linkage expressed as log of odds ratio to the null hypothesis (LOD score), Z score, or *p*-value
- Threshold statistics depend on study design and require correction for genome size (*n* hypotheses tested)

Genome-Wide Mapping

- Parametric analysis: known or assumed mode of inheritence
- Non-parametric analysis: all modes considered or mode independent

What is a LOD score?

- LOD is the log of the odds ratio of the probability of the data occurring under the specific hypothesis relative to the null hypothesis:

$$\text{LOD} = \log \left\{ \frac{\text{Probability assuming linkage}}{\text{Probability assuming } \underline{\text{no}} \text{ linkage}} \right\}$$

What is a LOD score?

- For multipoint analysis, the LOD is generally calculated for a recombination fraction, θ, from each marker
- Results are often reported as a maximum likelihood value: LOD (x) at θ =(y).

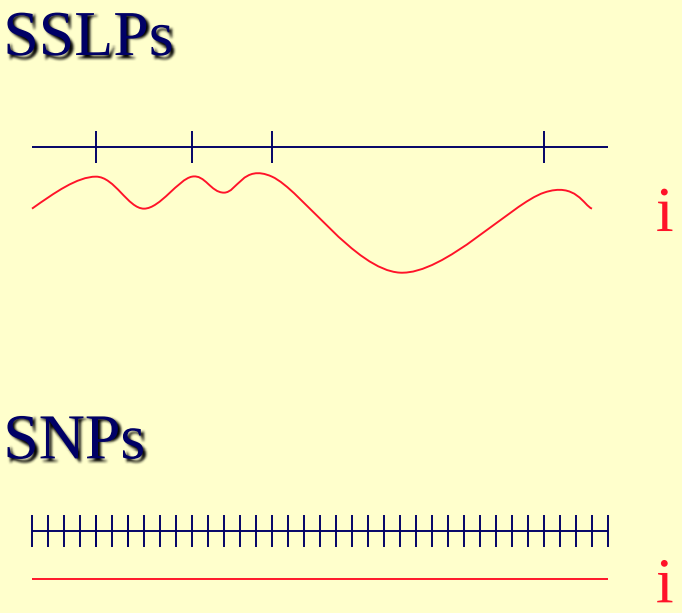
(hopefully with a high x and a low y)

Genome-Wide Markers

- **RFLPs**: few alleles, hard to detect
- **VNTRs**: many alleles, easier to detect
- **SSLPs**: many alleles, multiplex, high μ, ~1/20kb
- **SNPs**: biallelic, **mega**plex?, low μ, ~1/kb
- **Indels**: biallelic, **mega**plex?, low μ

Advantages of SNPs

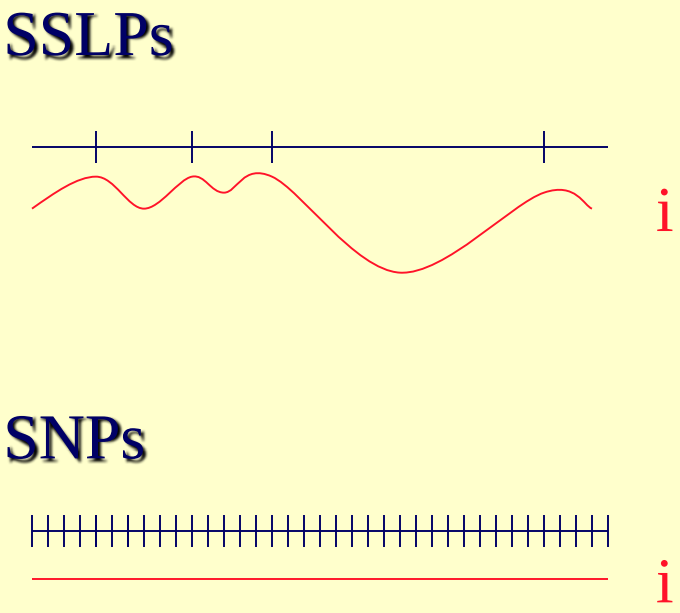
Biallelic markers create high information content **per haplotype** rather than **per marker**



Advantages of SNPs

- Mutation rate is negligible in pedigrees; corrected by haplotype in populations

- Non gel-based assays are easier to automate, multiplex and scale

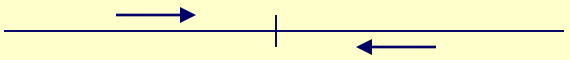


Detection of SNPs

- Allele-specific PCR
- Allele-specifc ligation
- Oligonucleotide hybridization (*Affymetrix, others*)
- TaqMan (*PE Biosystems*)
- Sequence tagging (ala SAGE)
- Single-Nucleotide Primer Extension (SNuPE)
  - fluorescent terminator (*PE Biosystems, AP-Biotech*)
  - mass tags for spectroscopy (*Sequanom, QiagenGenomics,*)

Detection of SNPs

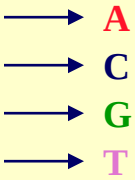
SNuPE
(minisequencing)



PCR

Extend with tagged ddNTP mix

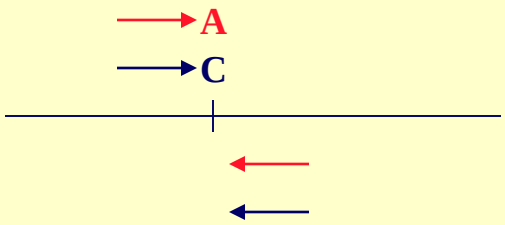
Capture product



Read out ratio of products

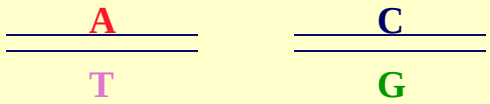
Detection of SNPs

Allele-Specific Ligation
(LCR assay)
for an A/C SNP



⇩ Ligase Chain Reaction

⇩ Capture product

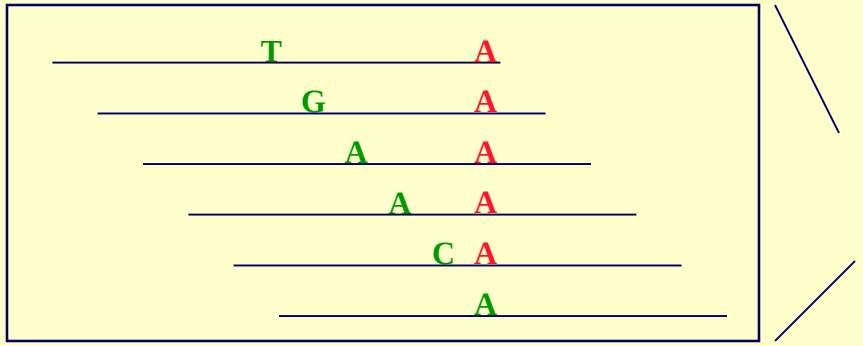


Read out ratio of products

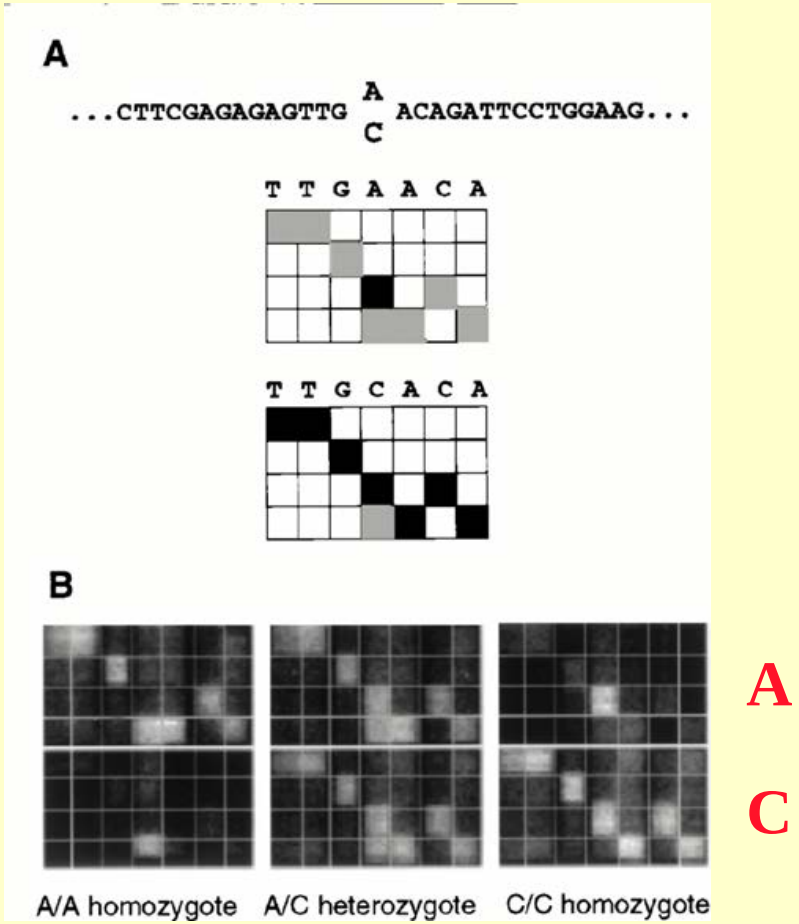
Detection of SNPs

Allele-specifc
Oligonucleotide
Hybridization

Tiled oligos query central base:



“A” tiling path



D.G.Wang et al. (1998) Science 280:1077-1082

Validation of SNP assays

Do allele calls cluster on unknown samples?