

# Genomics

BIT 220

Chapter 21

# Basic Terminology

Autosomes vs Sex Chromosomes

**Autosomal Recessive** need 2 copies of gene

Sickle-Cell Anemia

Cystic Fibrosis

**Autosomal Dominant** need 1 copy of gene

Huntington's Disease

**X-linked Dominant**

**X-linked Recessive**

Hemophilia

Color Blindness

## **Genomics**

mapping, sequencing and analyzing function of entire genomes (haploid set of chromosomes)

A. structural

nucleotide sequences

B. functional

transcriptome – complete set of RNAs

proteome – complete set of proteins

## **Proteomics**

structure and function of all proteins (and interactions)

## **Bioinformatics**

fusion of computer science and biology

analyzing and comparing different genomes

look for homologies – analyze structure and function

**SEE Figure 2- Technical sidelight (page 517)**

# MAPS Figure 21.1(p. 519)

## 1. Genetic (Linkage) Map

- linear array of genes on a chromosome based on recombination frequencies
- order of genes along a chromosome

## 2. Cytological

diagram of chromosome based on banding pattern

## 3. Physical Map

- actual distance in bp between two sites  
(ie restriction sites, sequence tagged sites)

# Genetic Map

- Exchanges of genetic information (crossing over, recombination) occur during meiosis
- Genes which are close to one another on a chromosome are typically linked together and inherited as a set
- The further away two genes lie from one another, the less likely they will be inherited together
- Recombination occurs with increasing frequency as the distance between two genes increases
- Use % of recombination to measure distance between genes
- This is **NOT** a precise physical distance!! – a good correlation, but recombination frequencies not always perfect

# Correlation between % Recombination and bp of DNA

1 map unit = 1 centimorgan (cM)

1 cM = 1% recombination =  $1 \times 10^6$  bp  
= 1 million base pairs

# Physical Map

## A. Contigs **Fig 21.6**

Construct a genomic DNA library use artificial chromosome

yeast (YAC)

bacterial (BAC)

bacteriophage (PAC)

Find 'DNA clones' from library that overlap forming a contiguous array (contig)

MAP a region or whole chromosome

## B. Restriction Maps

# Physical map (cont'd)

- **Anchor markers** – DNAs mapped both genetically and physically
- **STS maps** (sequence-tagged sites): short unique (single) copy segment of DNA; positioned by in situ hybridization
- **ESTs** (expressed-sequence tags): uses short cDNA sequences

# Polymorphic Site

A locus that has two or more alleles that occurs at a frequency of 1% in a population

## SNPs

single nucleotide polymorphisms

one base change between alleles

can effect protein or NOT – but can help identify map position

# Restriction Fragment Length Polymorphism **FIG 21.2**

Restriction site nullified by mutation or SNP

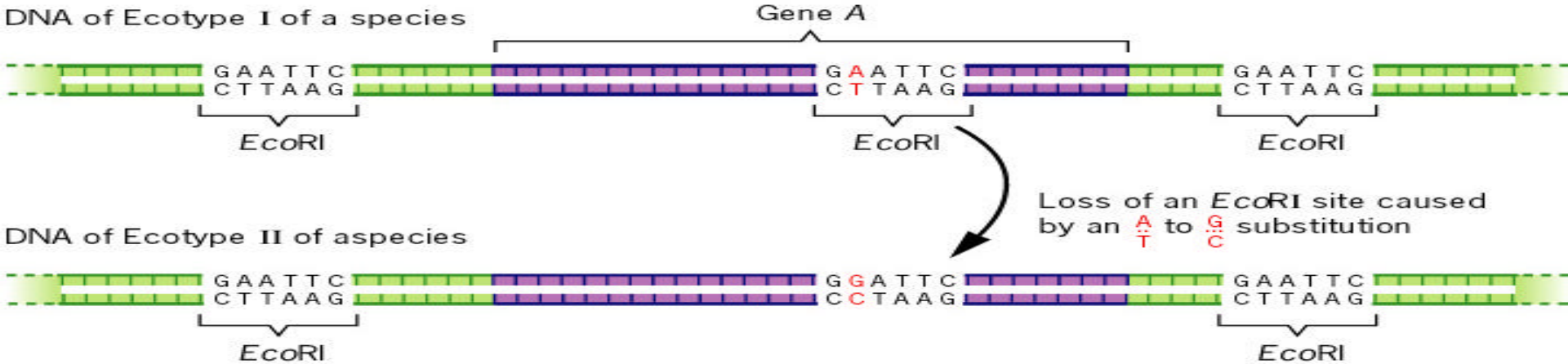
Different fragment sizes

Distinctive banding pattern

These polymorphic restriction endonuclease sites give us a marker loci for linkage mapping

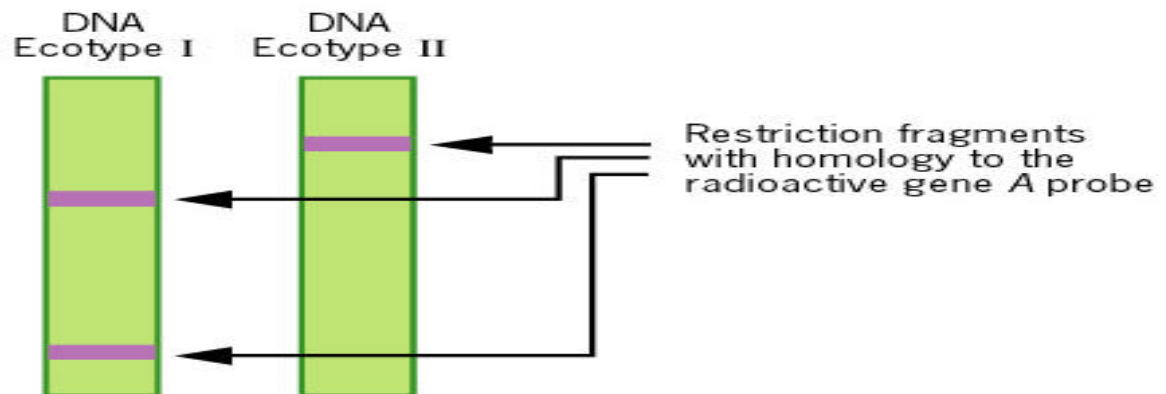
- no longer need to rely on phenotype

Help us to distinguish between geographical isolates, inbred species, individuals

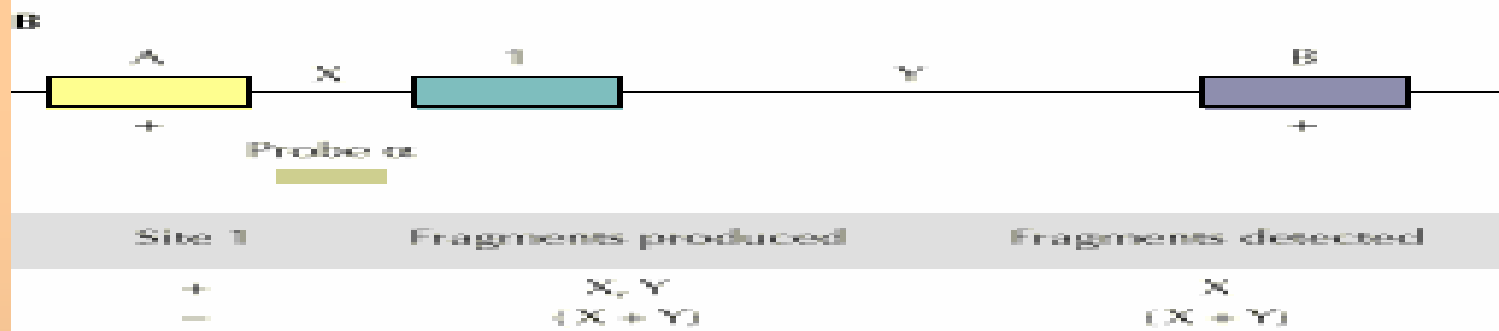


(a) Mutational origin of an RFLP

- STEP 1 Isolate DNA from each ecotype.
- STEP 2 Digest DNAs with restriction enzyme *EcoRI*.
- STEP 3 Separate DNA restriction fragments by agarose gel electrophoresis.
- STEP 4 Transfer DNA restriction fragments to nylon membrane.
- STEP 5 Hybridize DNA fragments on Southern blot to radioactive gene *A* clone.
- STEP 6 Wash blot and expose it to X-ray film to produce autoradiogram.



(b) Detection of an RFLP



# RFLP :

## Type 1 Short Tandem Repeat Polymorphisms

Short Tandem Repeats (Microsatellites)

Multicopy tandem repetitive sequences (2-3 bases)

Different number of repeats create different alleles

DNA from different individuals is PCR'd

The PCR products are run on gel

No polymorphism in STR locus of A (1 allele)

Polymorphism at STR locus of B (2 alleles)

# Minisatellite DNAs

Allele 1

(CA)<sub>15</sub>



CACACACACACACACACACACACACA  
GTGTG TG TG TGTGTGTGTGTGTGTGTGT



Unique  
DNA sequence

Unique  
DNA sequence

Allele 2

(CA)<sub>10</sub>



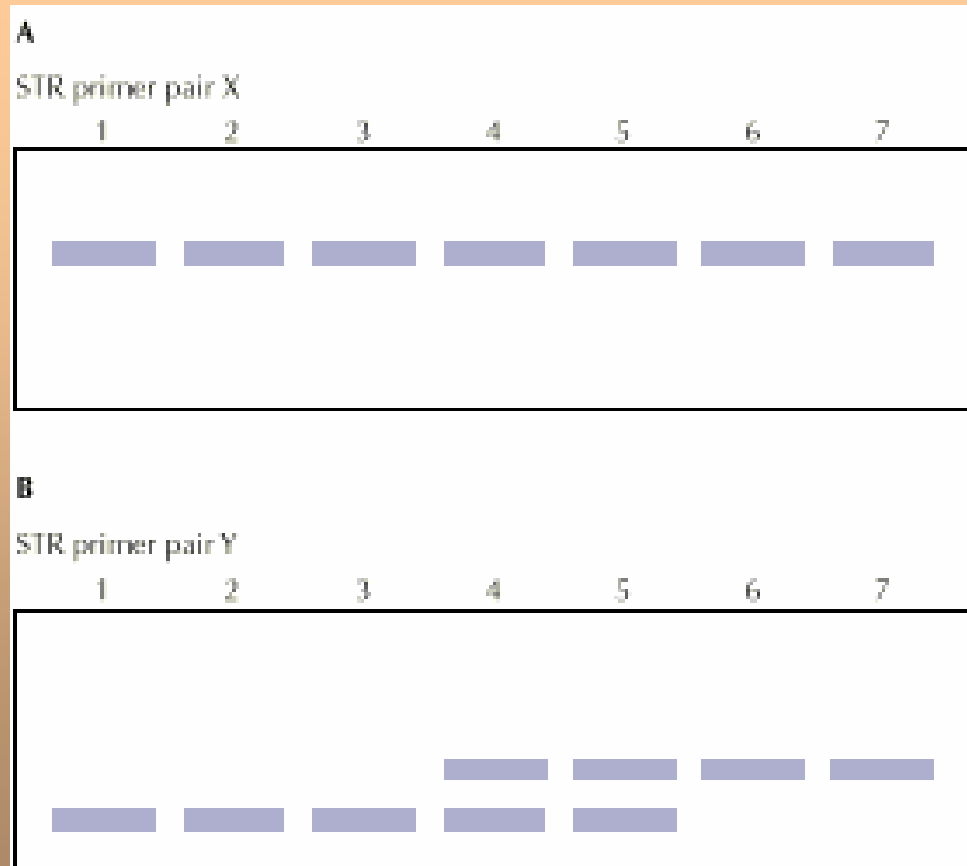
CACACACACACACACACA  
GTGTG TG TG TGTGTGTGTGTGTGT



Unique  
DNA sequence

Unique  
DNA sequence

# Analysis of Minisatellite DNA



# RFLP:

Type 2 Variable Number Tandem Repeats VNTRs

Different numbers of repeats between each restriction site

Vary in length from person to person (not in position of restriction

Sites but in number of repeats)

Very useful in human mapping

# Positional Cloning

## Figure 21.8

Chromosome Walks – moving along the Chromosome and cloning subsequent pieces (slide next)

Also known as positional cloning – search for a gene depends on its location along the chromosome



Starting clone

**STEP 1** Prepare a restriction map of the clone.



**STEP 2** Subclone B-H fragment.



**STEP 3** Screen genomic library with subclone as probe.



**STEP 4** Prepare a restriction map of the new clone.



**STEP 5** Subclone H-E fragment.



**STEP 6** Rescreen library with new subclone as probe.

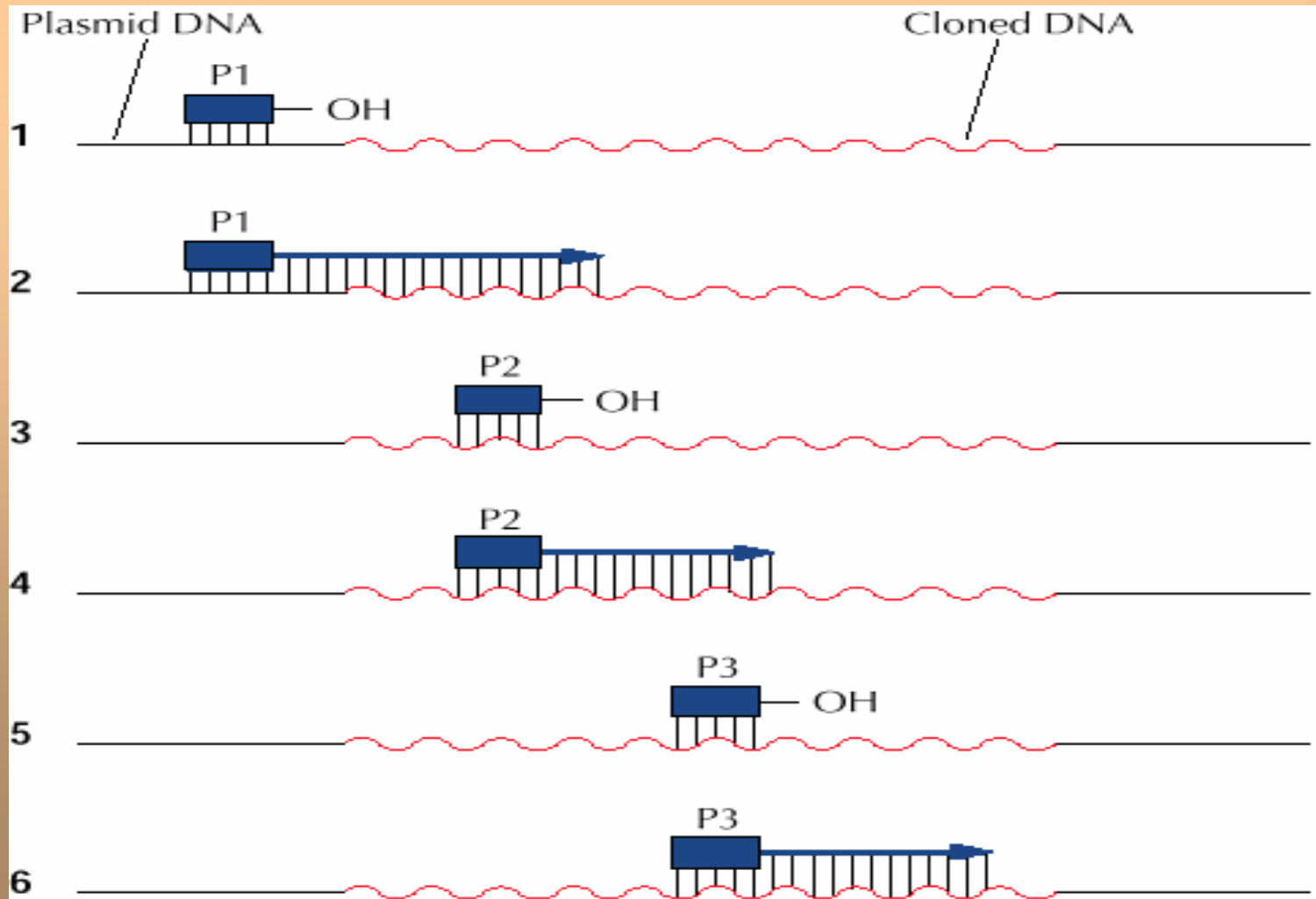


Repeat steps 1-3 for as many cycles as needed to reach the gene of interest

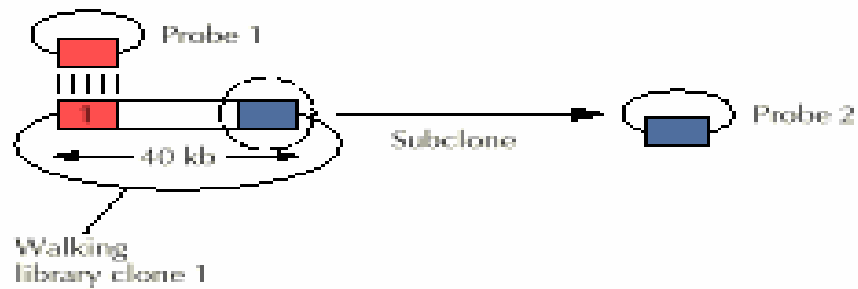
Clone of gene of interest



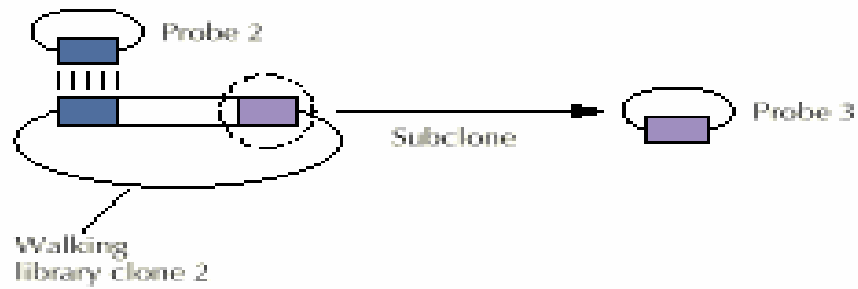
# Primer Walking



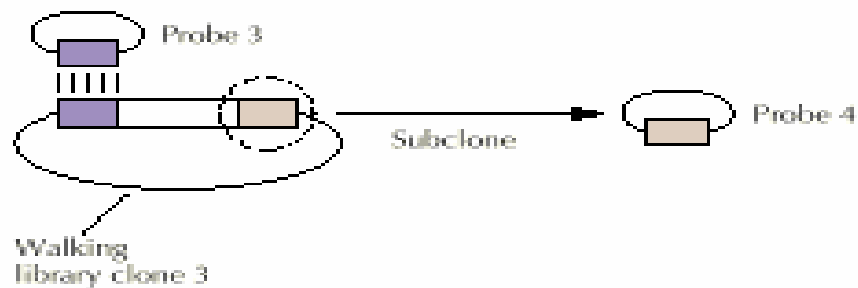
1



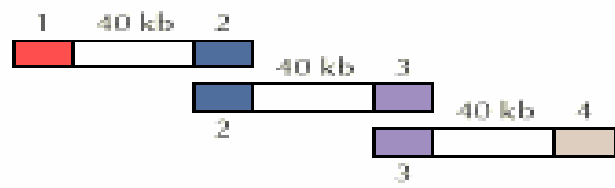
2



3



4 Extent of walk (from sequence 1 to sequence 4)



# Genomes sequenced

- Bacterial
- *Saccharomyces cerevisiae* (yeast)
- *Caenorhabditis elegans* (worm)
- *Drosophila melanogaster* (fruit fly)
- *Arabidopsis thaliana* (weed)
- Human

# Bacterial Genome

- 1<sup>st</sup> done – *Haemophilus influenzae* (1995)
- Since then, 32 bacterial genomes
- Range in size from 580,000 bp to 4.6 million bp (*E.coli*)
- 4288 putative protein sequences (have ORFs- open reading frames)

# *Saccharomyces cerevisiae*

- 1<sup>st</sup> eukaryotic – entire genome sequenced
- 12,068 kb (1996)
- Considerable genetic redundancy compared to bacterial genome- duplicate or multiple gene copies
- Distinguishing feature between eukaryotes vs. prokaryotes

# *Caenorhabditis elegans*

- Many developmental studies done with worm
- Simple genome – small genome (97mb-megabase pairs, or 97,000,000 base pairs)
- Little highly repetitive DNA sequences
- About 19,000 genes

# *Drosophila melanogaster*

- Model genetic organism of 20<sup>th</sup> century
- 180 mb genome
- Completed March 2000
- Only about 13,000 genes – less than worm

# *Arabidopsis thaliana*

- First plant to have genome sequenced
- Small genome: 125 mb
- Again, not much repetitive DNA
- About 25,000 genes
- About 70% of genes are duplicated – more than any other genome (so fewer than 15,000 unique gene sequences)

# Human Genome

- Human Genome Project original goals:
  - 1. map all the human genes
  - 2. construct physical maps of all 24 chromosomes (22 autosomes, X and Y)
  - 3. sequence entire genome by 2005

All goals ahead of schedule – first draft map published in 2001

# Human Genome

- First draft over 2650 mb (2,650,000,000 bp)
- 30,000 to 35,000 genes, rather than around 100,000 originally predicted
- Celera genomics – gives functional map of over 26,000 genes – **Figure 21.18**
- One gene every 60-85 kb in the genome

# Techniques for analysis

- “Gene chip” experiments (microarray hybridization experiments)
- <http://www.bio.davidson.edu/courses/genomics/chip/chipQ.html>
- **Figure 21.22 show genes are transcribed**
- Green fluorescent protein
- **Figure 21.23 show genes are translated**

# Other topics

- Proteomics – other power point (borrowed)
- [http://www.csupomona.edu/~drlivesay/Chm562/chm562\\_proteomics1.ppt](http://www.csupomona.edu/~drlivesay/Chm562/chm562_proteomics1.ppt)
- Pharmacogenetics (also called pharmacogenomics) – see Roche Genetics CD - The convergence of pharmacology and genetics dealing with genetically determined responses to drugs.