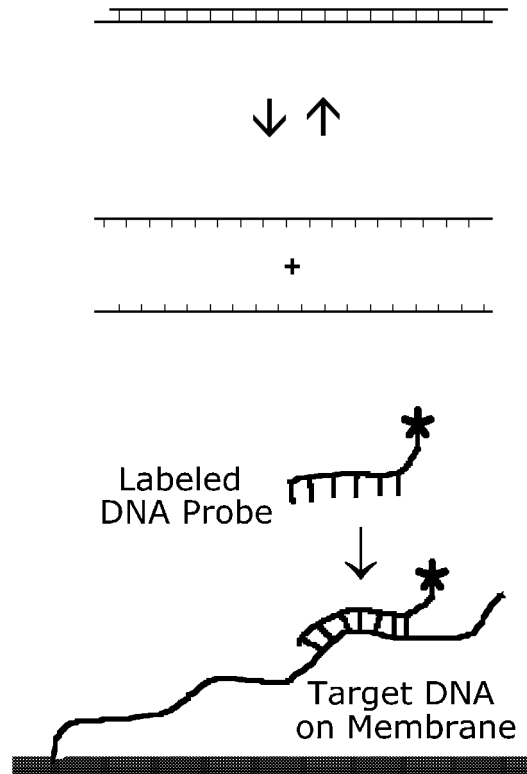
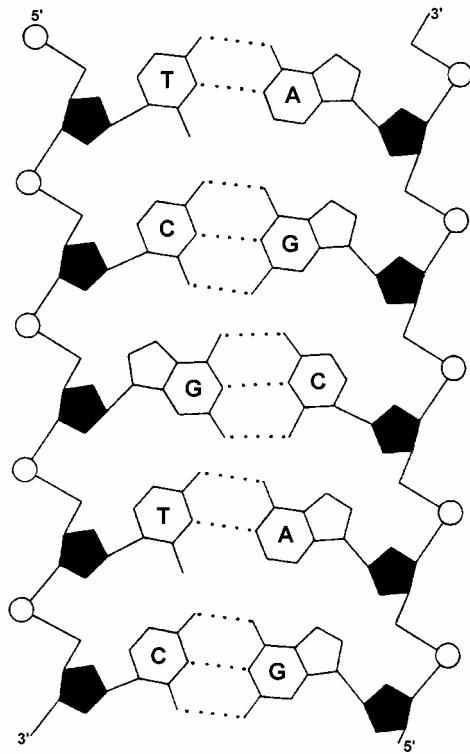


DNA Methods and Applications



Restriction Fragment Length Polymorphisms (RFLP)

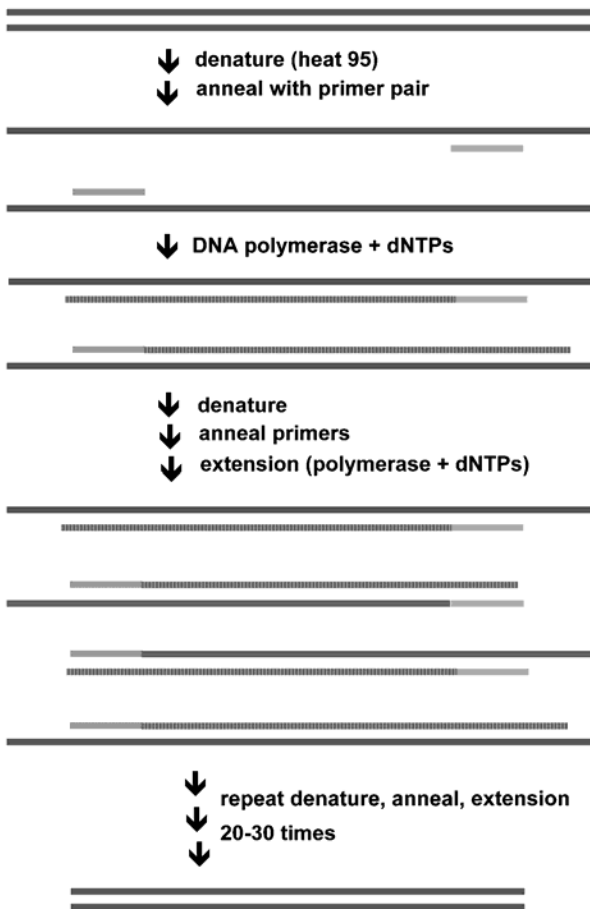
- digest DNA with restriction enzymes (cut DNA at specific sequences)
- separate fragments according to size by gel electrophoresis
- detect fragment(s)
- polymorphisms due to sequence differences

Genetic Fingerprinting

- some DNA probes produce complex RFLP patterns
- distinguish species, strains, individuals, etc.
- can be used in diagnosis, taxonomy, forensics, epidemiology, etc.

Polymerase Chain Reaction (PCR)

- specific DNA fragment(s) are enzymatically amplified
 - 10^6 -fold amplification possible
 - can detect single molecule
- tolerates impure DNA
- assay time < day
- possible to sequence PCR products



Molecular Epidemiology

- typing and subtyping of pathogens
 - genetic (eg., RFLP, PCR, sequence)
 - immunological (eg., antibodies, serology)
 - biochemical (eg., isoenzymes)
- 2 major types of analyses:
 - population genetics (current snapshot)
 - phylogenetic (dynamics and history)
- potential applications:
 - diagnosis
 - pathogenesis and virulence
 - drug resistance
 - transmission dynamics
 - taxonomy and evolution

DNA Sequence and Genomics

- tremendous amount of DNA sequence information
 - several entire genomes (eg., human, mosquito, and malaria parasite)
- potential applications
 - virulence factors
 - therapeutic targets
- huge need for bioinformatics (or computational biology)
 - sequence data management and analysis
 - searching databases

Databases

- two types: primary and secondary
 - 1^o contains original biological data (eg., DNA sequence)
 - 2^o has value added (eg., confirm protein sequence, structural data, etc)
- subdivisions (genome projects, taxonomic groups, etc)
- annotated to include ancillary information (keywords, gene name, author, publications, etc.)

Searching Databases

- text-based (annotations)
 - gene name, species, authors, etc.
- information retrieval systems
 - access all databases + medline
- sequence comparisons
 - submit 'query' sequence
 - compare to all sequences in database(s)
- similarities identified by aligning DNA or protein sequences
 - eg., BLAST

```
>pir||T24865 hypothetical protein T12D8.8 -Caenorhabditis elegans (Length = 422)
```

```
Score = 86.2 bits (210), Expect = 5e-16
Identities = 44/101 (43%), Positives = 60/101 (58%), Gaps = 2/101 (1%)
```

```
Query: 119 EAVDLVENKKYEEALEKYNKIISFGNPSAMIYTKRASILLNLKRPKACIRDCTEALNLNV 178
      +A + N ++ AL + I SAM++ KRA++LL LKRP A I DC +A+++N
Sbjct: 121 KAQEAFSNGDFDTALHTFTAAIEANPGSAMLHAKRANVLLKLRPVAAIADCDKAISINP 180
```

```
Query: 179 DSANAYKIRAKAYRYLGKWEFAHADMEQGQKIDYDE--NLW 217
      DSA YK R +A R LGKW A D+ K+DYDE N W
Sbjct: 181 DSAQGYKFRGRANRLLGKWVEAKTDLATAACKLDYDEAANEW 221
```

Gaps to maximize alignment

Probability of match appearing by chance

```
Query= Pbpp58b (423 letters)
```

```
Database: nr (493,611 sequences; 154,780,071 total letters)
```

Sequences producing significant alignments:	Score (bits)	E Value
sp Q08168 HRP_PLABE 58 KD PHOSPHOPROTEIN (HEAT SHOCK-RELATED PRO...	334	1e-90
gb AAC37300.1 (L21710) 58 kDa phosphoprotein [Plasmodium berghei]	329	3e-89
pir T10455 heat shock related protein - Plasmodium berghei >gi ...	250	2e-65
sp P50503 HIP_RAT HSC70-INTERACTING PROTEIN >gi 4379408 emb CAA5...	106	5e-22
sp P50502 HIP_HUMAN HSC70-INTERACTING PROTEIN (PROGESTERONE RECE...	87	3e-16
gb AAF45894.1 (AE003429) CG2947 gene product [Drosophila melano...	87	4e-16
pir T24865 hypothetical protein T12D8.8 - Caenorhabditis elegans...	86	5e-16
pir T04562 hypothetical protein T12H17.60 - Arabidopsis thalian...	81	2e-14

Identifier Line:
database | accession # | name or locus

Alignment Score

Genomes and Proteomes

- genome: complete set of genetic information in organism
- proteome: complete set of proteins in cell, tissue, or organism

