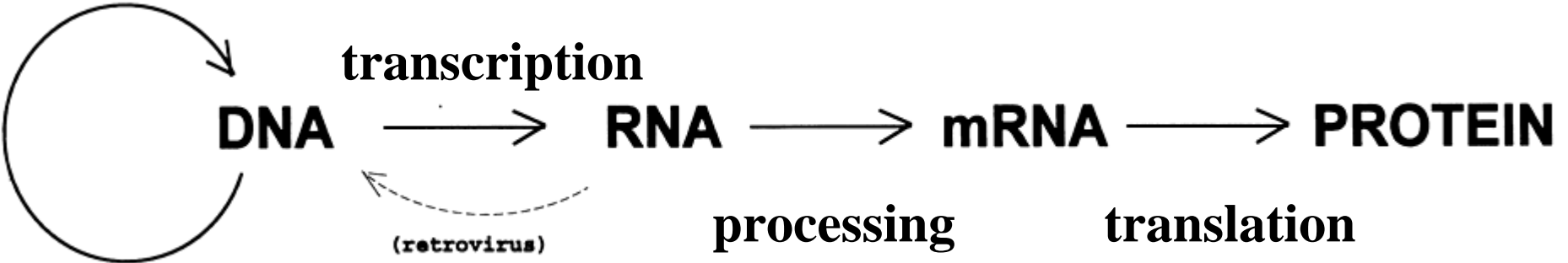
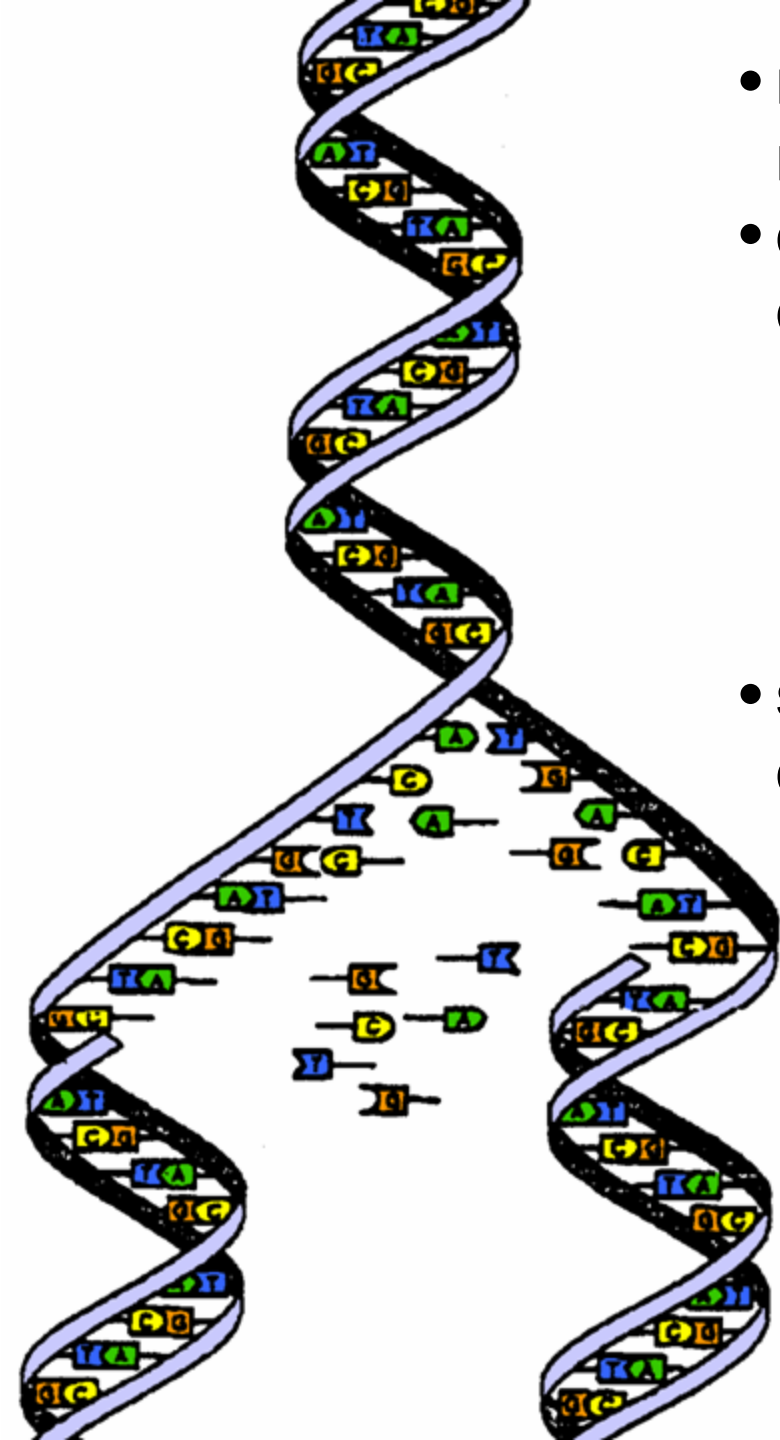


# Molecular Analysis

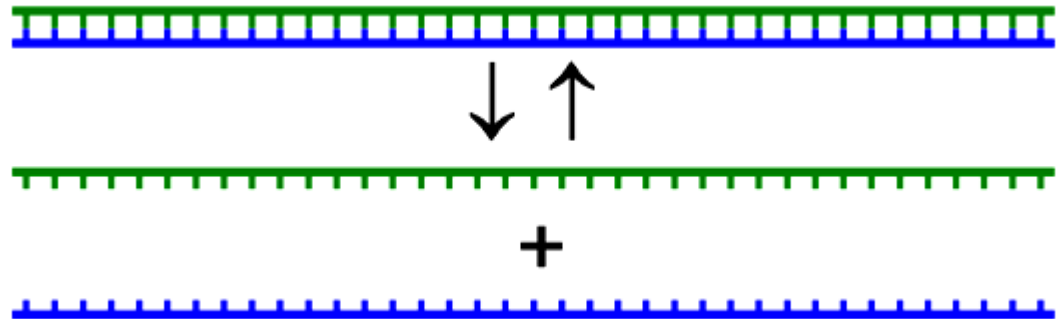
replication



- possible to detect and analyze DNA, RNA, and protein
- DNA sequence represents 'genotype'
- expressed genes (ie, mRNA and protein) represent 'phenotype'
- generally analysis of DNA is easier than RNA or protein

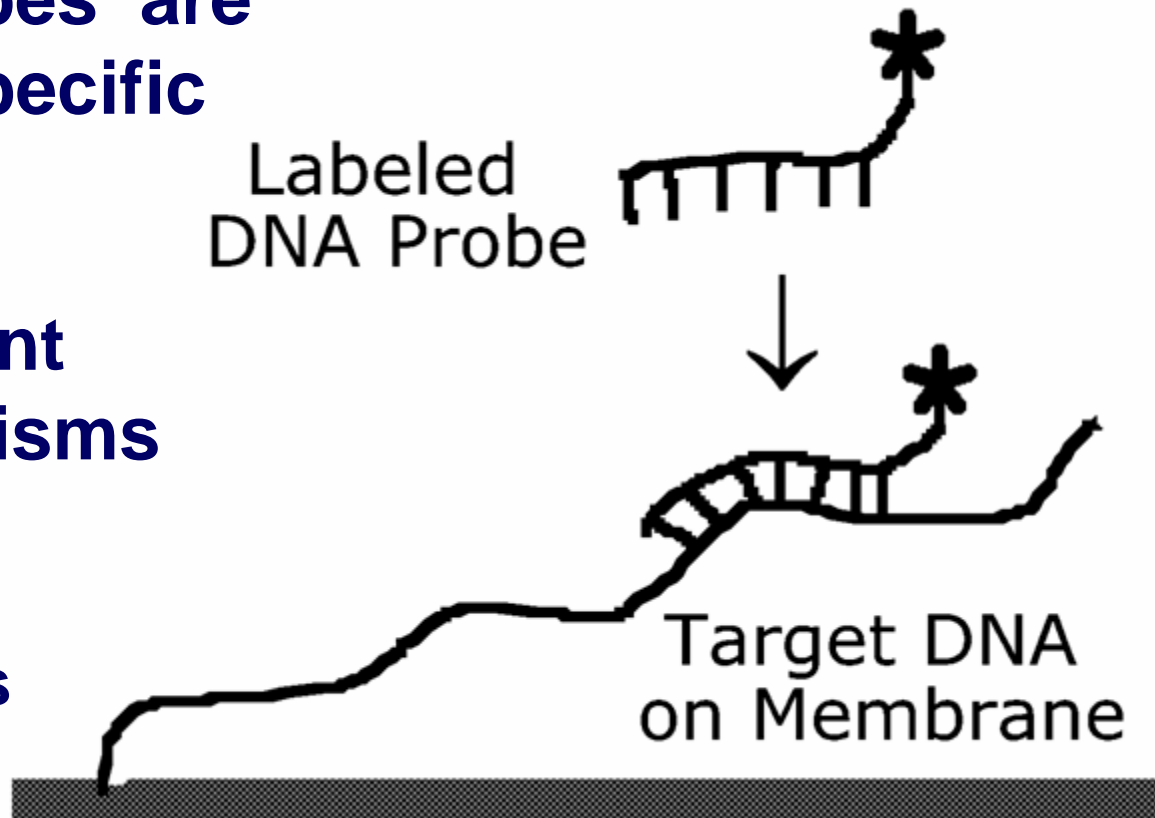


- nucleic acids are polymers of nucleotides
- double-stranded DNA composed of complementary strands
  - hydrogen bonds (weak)
  - determined by specific base pairing (A:T and G:C)
  - template for the synthesis
- specific base-pairing permits detection and analysis of DNA/RNA



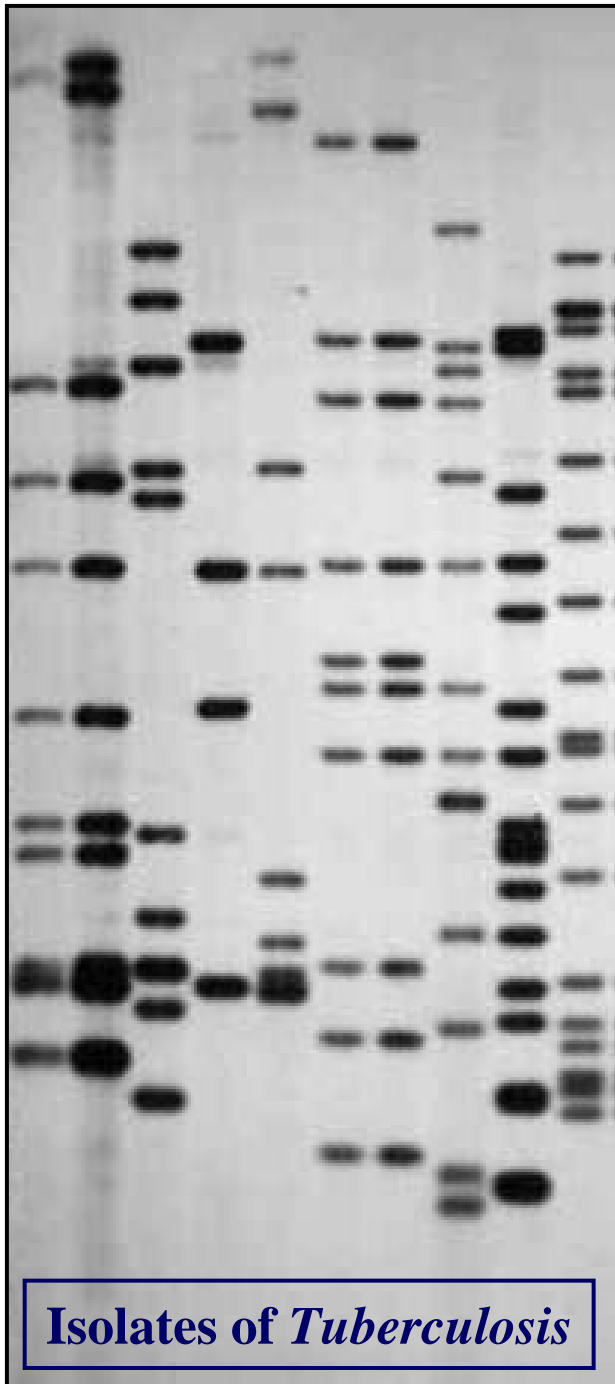
# Specific Nucleic Acids Can Be Detected with Probes

- homologous 'probes' are used to detect specific nucleic acids
- restriction fragment length polymorphisms (RFLP)
  - digest DNA
  - separate fragments
  - detect fragments



# Genetic 'Fingerprinting'

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



*Isolates of Tuberculosis*

## RFLP Limitations

- needs  $\mu\text{g}$  amounts of DNA
- DNA needs to be pure
- assay time: several days to > week

# **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**

↓ denature (heat 95)  
↓ anneal with primer pair

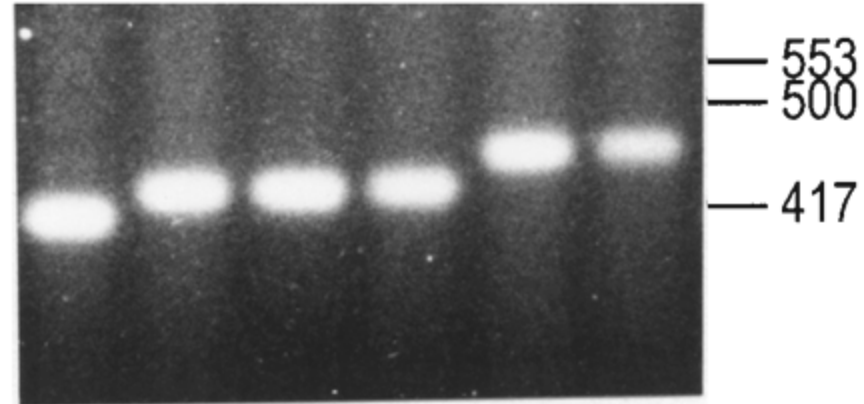
↓ DNA polymerase + dNTPs

↓ denature  
↓ anneal primers  
↓ extension (polymerase + dNTPs)

↓ repeat denature, anneal, extension  
↓ 20-30 times

## PCR Requirements

- target DNA
- oligonucleotide primers
- *Taq* polymerase + dNTPs
- thermocycler



- ↗
- analyze PCR products by gel electrophoresis
  - and/or sequence

# 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° contains original biological data (eg., DNA sequence)
  - 2° 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

# Example of Blast Search

Score = 86.2 bits (210), Expect = 5e-16

Identities = 44/101 (43%), Positives = 60/101 (58%), Gaps = 2/101 (1%)

```
Query: 119 EAVDLVENKKYEEALEKYNKIISFGNPSAMIYTKRASILLNLKRPKACIRDCTEALNINV 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
```

Similarity

Gap

Matches

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

database | accession # | entry name or locus

Probability

# Genomes and Proteomes

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

