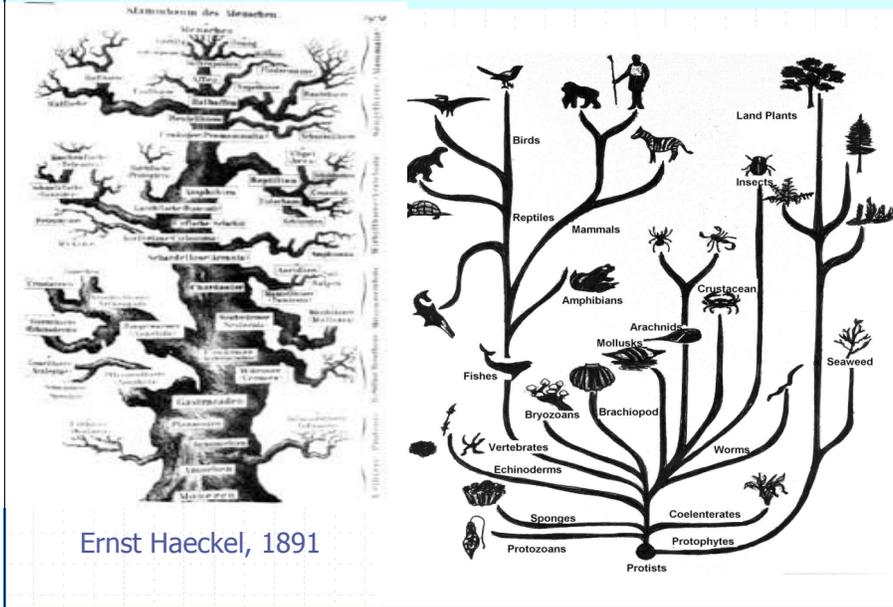# Chapter 3: Phylogenetics

3.1 Introduction To Evolution

Prof. Yechiam Yemini (YY)
Computer Science Department
**Columbia University**

Overview

- Introduction to evolution
- Markovian models of evolution
- Computing Phylogeny
  - Distance-based techniques
  - Maximal parsimony (character based) techniques
  - Probabilistic approaches
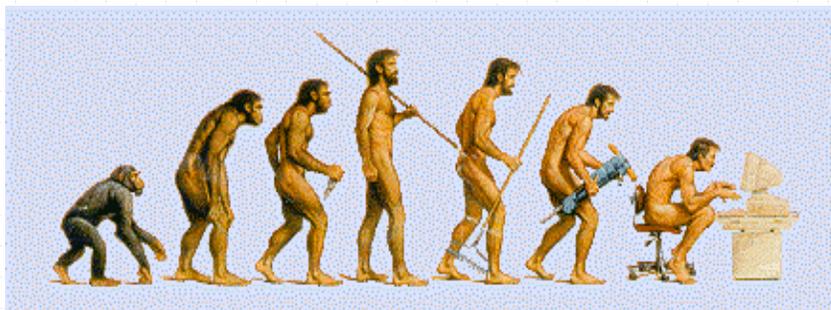
The Tree of Life

Ernst Haeckel, 1891

Traditional Approaches

- Phenetic: taxonomy of phenotypical similarity

- Cladistic: tree of likely evolutionary pathways [W. Henning, 1950]
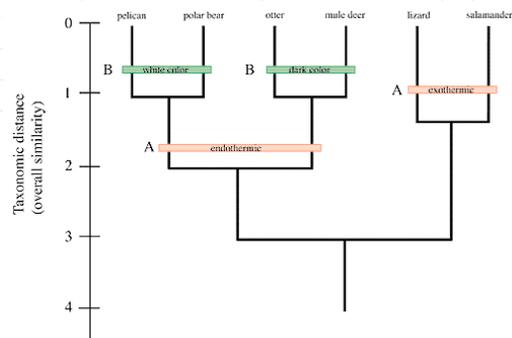
Phylogenetics: Computing Evolution

■Given: related taxa of species
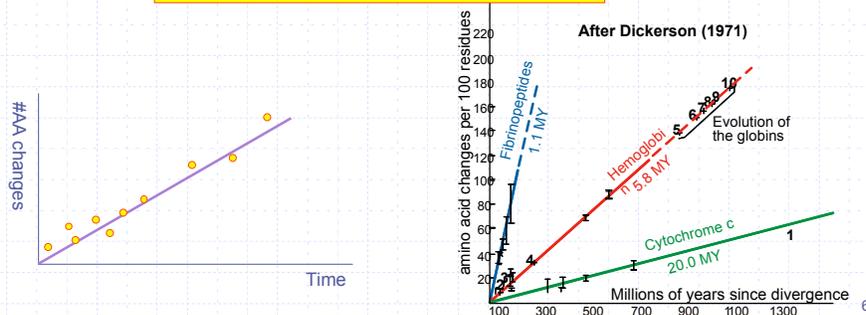
■Compute: evolutionary tree of these taxa

5

Molecular Evolution

Zuckerland & Pauling 1962!!!
- Closely related species have fewer amino acid differences
- AA changes are related to the time of divergence
- The molecular clock hypothesis
  - Confirmed by Fitch [62], Kimura[83]….
  - Central role in molecular phylogenetic

Molecular Clock Hypothesis:
AA Changes Occur At A Fixed Rate
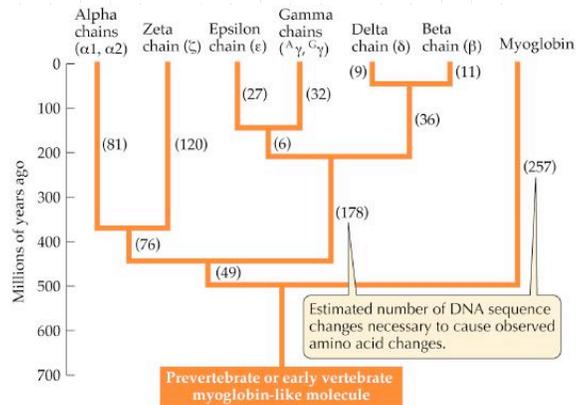

After Dickerson (1971)

Fibrinopeptides 1.1 MY

Hemoglobin 5.8 MY

Evolution of the globins

Cytochrome c 20.0 MY

amino acid changes per 100 residues

Millions of years since divergence

6

#AA changes

Time

3

Molecular Phylogenetics

- Given: evolutionary related sequences (AA/DNA)
- Compute: phylogenetic tree of the sequences
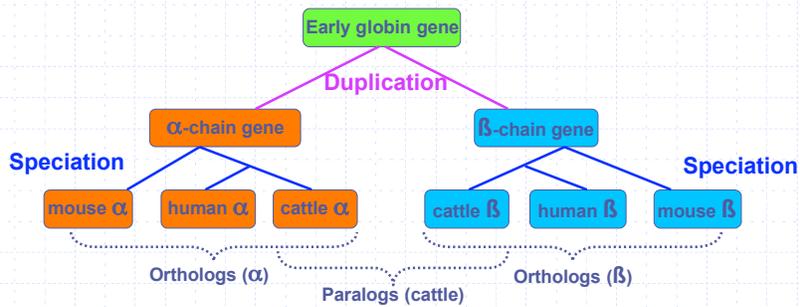
Evolution Tree of Globins

© 1998 Sinauer Associates, Inc.

7

Pathways Of Molecular Evolution

- Basic mechanisms:
  - Gene speciation; duplication; transfer [e.g., virus]
- Similar genes can be
  - Orthologs: have common ancestor
  - Paralogs: duplicates of same gene
  - Xenologs: formed through transfer of genes

8

Where Do Panda's Come From? [Pevzner]

■Resolve through DNA phylogeny [O'Brien et al 1985]

BROWN BEAR POLAR BEAR BLACK BEAR SPECTACLED BEAR GIANT PANDA RACCOON RED PANDA
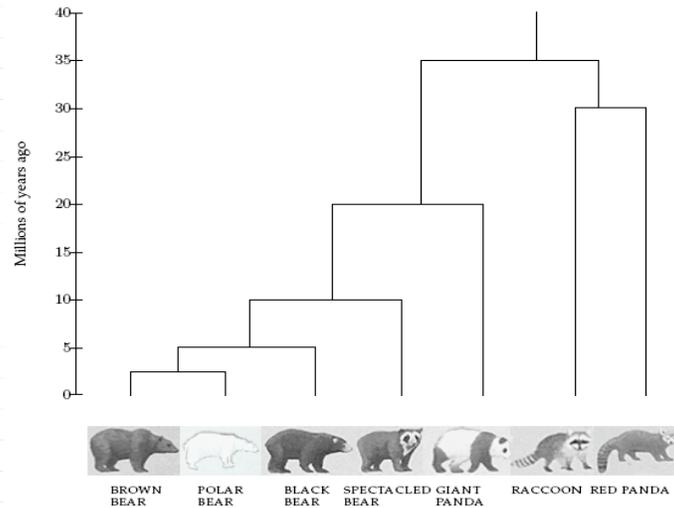
9

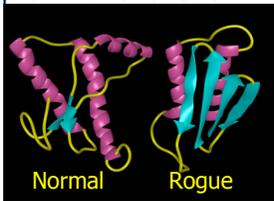# Analyzing Human Evolution
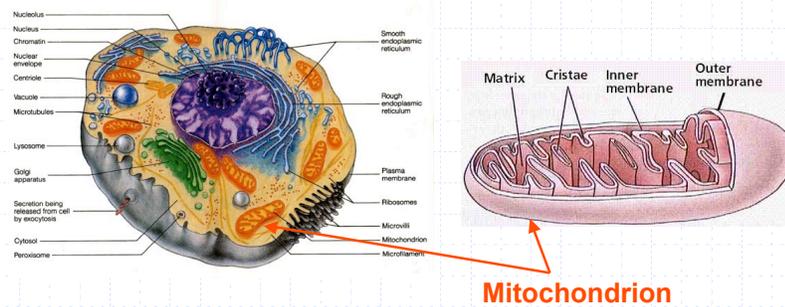
Mitochondrion

■ Mitochondrion
- Generator of ATP (energy store)
- mtDNA has 13 genes
- Transferred through maternal eggs

■ Compute phylogeny of mtDNA

10

5

Human Evolution

The application of molecular genetic approaches to the study of human evolution
Cavalli-Sforza & Feldman

Nature Genetics **33**, 266 - 275 (2003)

11

Human Evolution   www.sanger.ac.uk

Evolution of Prions
http://www.mad-cow.org

- Prions are brain proteins
- Can cause brain diseases
- What is their function?
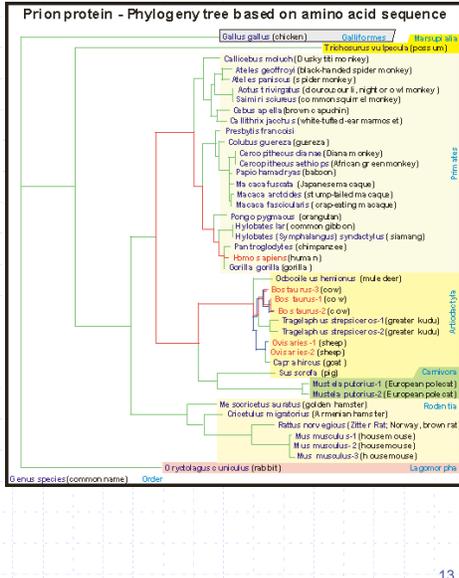 - Forming synaptic connections
    [E. Kandel, S. Kausik, 2004]

Normal   Rogue

Exon 2 Sequences Aligned [Clustal W] and % Identity

```
U29186 mouse    GACTCCTGAGTATATTTCAGAACTGAACCATTTCAACCGAGCTGAAGCAT 50
D50092 rat      GACTCCTGAATATATTTCAAAACTGAACCATTTCAACCCAACTGAAGTAT 50
U78769 hamster  GACTCCTGAATATATTTCCAAAACTGAACAATTTCAACTGACCTGAAGTAC 50
Consensus       GACTCCTGAATATATTTCAAAACTGAACAATTTCAACCAAGCTGAAGCAT 50
D26150 cow      GACTTCTGAATATATTTGAAAACTGAACAGTTTCAACCAAGCCGAAGCAT 50
U67922 sheep    GACTTCTGAATATATTTGAAAACTGAACAGTTTCAACCAAGCTGAAGCAT 50
U29185 human    GACTCCTGAATATTTTTCAAAACTGAACAATTTCAGCCATGTCTGAGCTT 50
                ****.****.***.**..*.********..*****.* .**
                        51 Conserved Residues

U29186 mouse    TCTGCCTTCCTAGTGGTACCAGTCCAATTT-AGGAGAGCCA-AGCAGACT 98
D50092 rat      TCTGCCTTCTTAGCGGTACCAGTCCGGTTT-AGGAGAGCCA-AGCCGACT 98
U78769 hamster  TCTGTTTTTCTTAGAGGTACCAGTTCAGTTT-AGGAGAGTCACAGCAGATC 99
Consensus       TCTGTCTTCCTAGAGGTACCAGTCCAGTTT-AGGAGAGCCACAGCAGATT 99
D26150 cow      -CTGTCTTCCCAGAGACACAAATCCAACTTGAGCTGAATCACAGCAGAT- 98
U67922 sheep    -CTGTCTTCCCAGAGACACAGATCCAACTTGAGCTGAATCACAGCAGAT- 98
U29185 human    TCCGTCTTCCTGGAGGCACAAATCTAGTTT-AGCTGAACCACAAACAGATT 99
                * * .**  * * .**.*    .** **.**. ** * **
```

13

Evolution of Prions
http://www.mad-cow.org

- Prions are brain proteins
- Can cause brain diseases
- What is their function?
  - Forming synaptic connections
    [E. Kandel, S. Kausik, 2004]

Normal   Rogue

Exon 2 Sequences Aligned [Clustal W] and % Identity

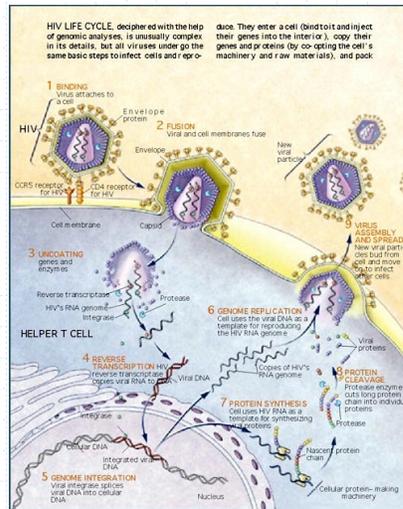(sequence alignment figure)

13

HIV Evolution
http://hiv-web.lanl.gov/

- HIV is an RNA virus
  - Hijacks cell mechanisms
  - Disables cell protections
  - Mutates rapidly
- Phylogenetic analysis is of central importance
  - Analyze HIV mechanisms
    → plan drugs
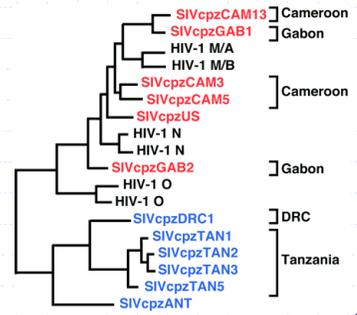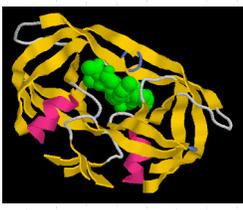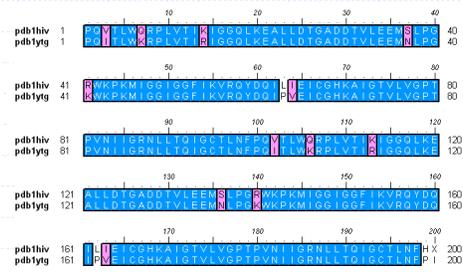 - Analyze HIV mutant strains and epidemic spread

http://www.cdc.gov/ncidod/EID/vol10no7/03-0904-G1.htm

14

HIV Evolution

- Tracing HIV evolution through protease homology

Sample Phylogenetic Analysis of HIV

http://www.mathworks.com/products/demos/bioinfo/hivdemo/hivdemo.html

- Step 1: extract gene sequences from GeneBank
  - Get coding sequences for key proteins [e.g., gag/pol/env]

'HIV-1 (Zaire)'	'K03454' [1 2 8] ;
'HIV1-NDK (Zaire)'	'M27323' [1 2 8] ;
'HIV-2 (Senegal)`	'M15390' [1 2 8] ;
'SIVMM251 Macaque'	'M19499' [1 2 8];

  ....

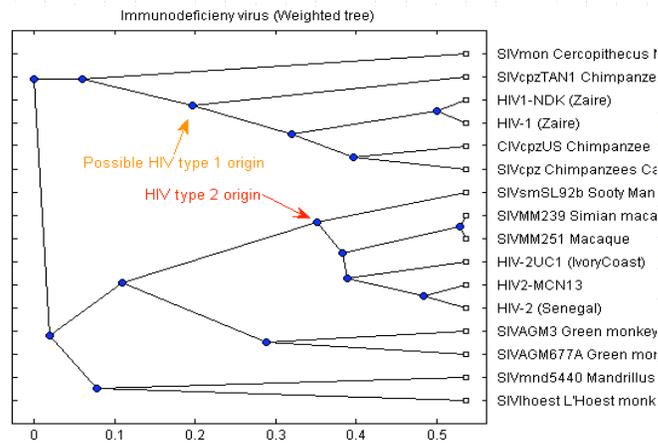- Construct evolutionary tree for each protein
  - Convert DNA sequence to protein sequence
  - Use UPGMA to construct tree

8
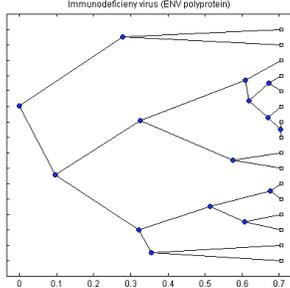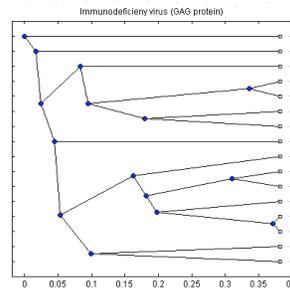
HIV Phylogeny

■ Build consensus tree (by averaging weights)
■ Use tree to analyze HIV origins and evolutionary spread

Immunodeficieny virus (Weighted tree)

Possible HIV type 1 origin

HIV type 2 origin

SIVmon Cercopithecus I
SIVcpzTAN1 Chimpanze
HIV1-NDK (Zaire)
HIV-1 (Zaire)
CIVcpzUS Chimpanzee
SIVcpz Chimpanzees Ca
SIVsmSL92b Sooty Man
SIVMM239 Simian maca
SIVMM251 Macaque
HIV-2UC1 (IvoryCoast)
HIV2-MCN13
HIV-2 (Senegal)
SIVAGM3 Green monkey
SIVAGM677A Green mor
SIVmnd5440 Mandrillus
SIVlhoest L'Hoest monk

0   0.1   0.2   0.3   0.4   0.5

17

Notes

■ Phylogenetics helps resolve central questions:
  • What mutations transform birds-flu to virulent human flu?
  • Which proteins are the best drug targets?
  • What mutations are responsible for a given hereditary disease?
  • ….

18

9