1/19

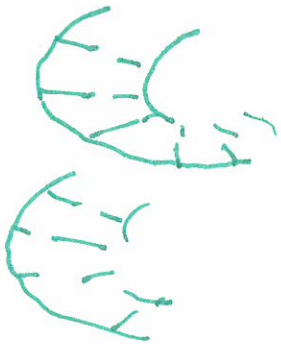
- Hand in HW.

Hand in HW.

- Announcements
- Trees
- Technical Lecture ✓

Basics of DNA.

2/28/1953 The Eagle in Cambridge, Eng.



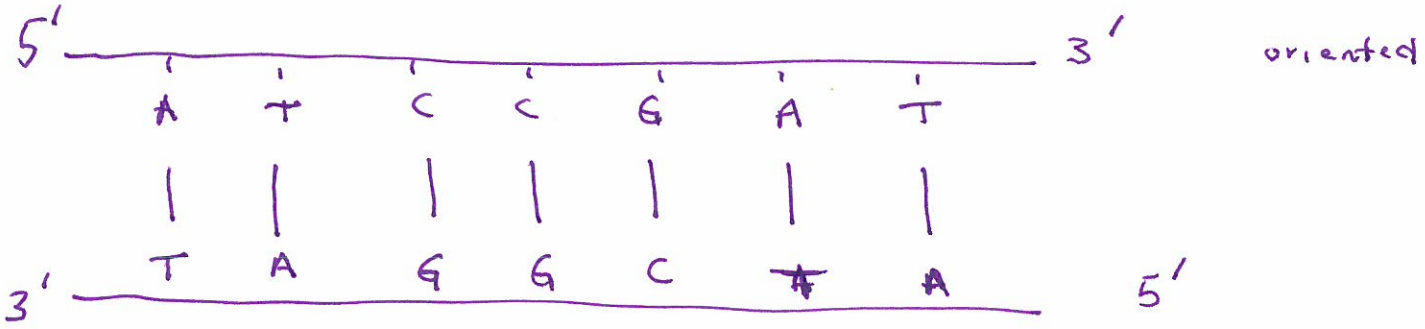
Watson and Crick
double helix

DNA made up of NUCLEOTIDES or BASES

AG — Purines — R — 2 rings

CT — Pyrmidines — Y — 1 ring

2-stranded DNA



base pairing $A \longleftrightarrow T$

$C \longleftrightarrow G$

Abstraction: A single strand of DNA is a SEQUENCE of 4 BASES

1	2	3	4	5	6	7
A	T	T	C	G	T	T

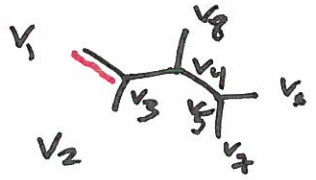
For phylogenetics, aligned sequences need to be

ORTHOLOGOUS ≡ descendent from a common ancestral base

TREES as mathematical objects in phylogenetics

Defn: An <u>unrooted tree</u> $T = (V, E)$ is a

connected graph with no cycles.



$T = (V, E)$ $\quad V = \{v_1, v_2, \ldots, v_8\}$ = vertices

E = edges = $\{ \{v_1, v_3\}, \{v_2, v_3\}, \ldots \}$

For vertices, there are INTERIOR vertices and LEAVES.
(= tips)

One notion of distance on a tree is called

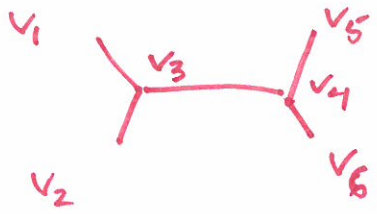
GRAPH-THEORETICAL DISTANCE.

This pairwise distance counts the number of edges

between two vertices.

$$d(v_1, v_2) = 2 \qquad d(v_3, v_7) = 3$$

$$d(v_6, v_6) = 0$$



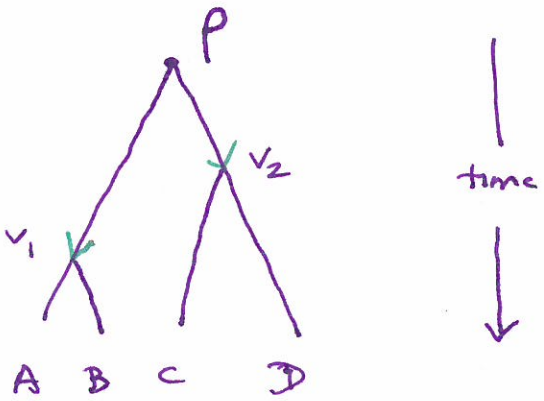
T

DISTANCE MATRIX

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	0	2	1	2	3	3
v_2		0	1	2	3	3
v_3			0			
v_4						
v_5						

upper diagonal

Biologists prefer rooted trees.



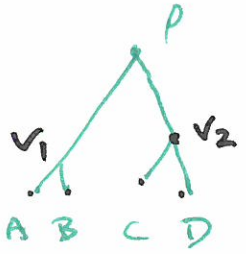
p = root

A tree with DIRECTED edges

$T = (V, E)$

$E = \{ (p, v_1), (v_1, A), (v_2, C),$ ... $\}$

parentheses



The root ρ is considered the

MOST RECENT COMMON ANCESTOR

MRCA.

Related to idea that sequences are ORTHOLOGOUS

For any vertex v, its VALENCE is the number of edges meeting at v.

$$Val(\rho) = 2 \qquad Val(v_1) = 3$$

v is a LEAF if its $Val(v) = 1$.

Defn: An unrooted tree T is BINARY if all interior vertices have valence 3.

A rooted " " " " " " "

" " " " "

∧

except the root.