Week 6 Phylogenomics

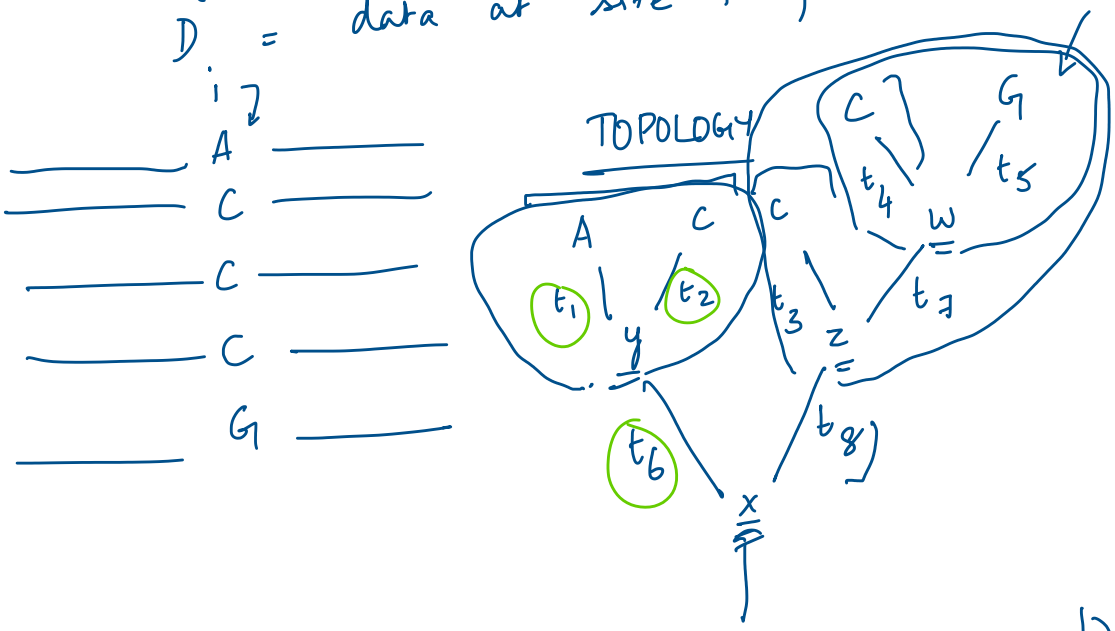
Monday, 28 February 2022 7:46 AM

ASSUMPTIONS

1. independent evolution at every site
2. independent evolution on every branch

$$L = P(D|T) = \prod_{i=1}^{m} P(D^{(i)}|T)$$

$D^{(i)}$ = data at site i, T = tree



$$P(D^{(i)}|T) = \sum_{x}\sum_{y}\sum_{z}\sum_{w} P(A,C,C,C,G,x,y,z,w|T) =$$

$x, y, z, w = \{A, C, G, T\}$

$$= P(x) \times P(y|x,t_6) \times P(A|y,t_1) \times P(C|y,t_2) \times P(z|x,t_8) \times P(C|z,t_3) \times P(w|z,t_7) \times P(C|w,t_4) \times P(G|w,t_5) =$$

$$(i) \ldots = P(x) P(y|x,t_6) P(A|y,t_1) P(C|y,t_2) \ldots$$

$$P(D^{(i)} \mid T) = \sum_x \Big(\sum_y P(y \mid x, t_6) \cdots \Big) \times \Big(\sum_z P(z \mid x, t_8) \times P(C \mid z, t_3) \Big) \times \Big(\sum_w P(w \mid z, t_7) \; P(C \mid w, t_4) \; P(G \mid w, t_5) \Big)$$

eg. $P(C \mid w, t_4) \times P(G \mid w, t_5)$ $\quad = \quad =$

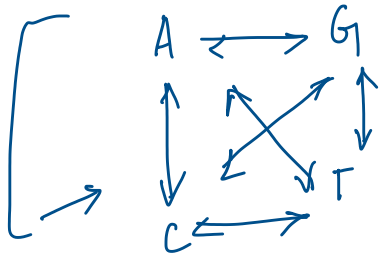
MUTATION MODELS

JUKES - CANTOR (1969)

each nucleotide has same (equal) prob. of changing into another.

TRANSITION PROB.

$$\pi_A = \pi_G = \pi_C = \pi_T = \frac{1}{4}$$



u = mutation rate

model mutation as a Poisson process on a branch; time until first event occurs

$$\Big[P(A \to A) + P(A \to C) + P(A \to G) + P(A \to T) = \frac{u}{3} + \frac{u}{3} + \frac{u}{3} + \frac{u}{3} \Big]$$

$$- 4u \;\Big]$$

$-\frac{4}{3}\mu t$]

⇒ Prob. of no event on branch of length 't' (time)

$= e^{-\frac{4}{3}\mu t}$

∴ P (at least one event) $= 1 - e^{-\frac{4}{3}\mu t}$

---

eg. $p(C|A,u,t) = \frac{1}{4}\left(1 - e^{-\frac{4}{3}\mu t}\right)$

$p(C|G,u,t) = \frac{1}{4}\left(1 - e^{-\frac{4}{3}\mu t}\right)$

...

KIMURA 2 Parameter model

$A \leftrightarrow G$, $C \leftrightarrow T$ α

$A \leftrightarrow C$, $G \leftrightarrow C$, $T \leftrightarrow A$, $T \leftrightarrow G$ β

Q MATRIX

	A	C	G	T
A	$-\mu(a\pi_C + b\pi_G + c\pi_T)$	$\mu a \pi_C$	$\mu b \pi_G$	$\mu c \pi_T$
C	$\mu g \pi_A$	$-\mu(g\pi_A + d\pi_G + e\pi_T)$	$\mu d \pi_G$	$\mu e \pi_T$

G
T
$\left[\quad \cdots \quad \right]$

μ = mutation rate

$a, b, c \ldots$ = rate parameters.

π_A = frequencies of A, C, G, T