Week 5 Phylogenomics

Friday, 18 February 2022 8:02 AM

African ancestry (American SW)

20x coverage LCT ; rs 4988235 SNP

70% G (ref) ; 30% A (alternate)

(1) Compute genotype likelihoods.

(2) Which is the most likely genotype at this locus?

TRUTH | GENOTYPE LIKELIHOOD

GG $\quad {}^{n}C_{k}\, \epsilon^{k} (1-\epsilon)^{n-k} = P(D|GG)$

AA $\quad {}^{n}C_{n-k}\, \epsilon^{n-k} (1-\epsilon)^{k} = P(D|AA)$

AG $\quad {}^{n}C_{k} \left(\frac{1}{2^n}\right) = P(D|AG)$

n = total # of reads (20)

$k = 14$

(1) $P(D|GG) = {}^{20}C_{6} \left(\frac{6}{20}\right)^{6} \left(\frac{14}{20}\right)^{14} = 0.191639$

(2) $P(D|AA) = {}^{20}C_{14} \left(\frac{14}{20}\right)^{6} \left(\frac{6}{20}\right)^{14} = 0.002181$

(3) $P(D|AG) = {}^{n}C_{k} \left(\frac{1}{2^n}\right) = {}^{20}C_{14} \left(\frac{1}{2^{20}}\right) = 0.0369$

$P(GG|D) \propto P(D|GG) \times P(GG)$

$\propto 0.191639 \times 0.705 = 0.135$

$P(AA|D) \propto P(D|AA) \times P(AA)$

$\propto 0.002181 \times 0.049 = 0.000106$

$P(AG|D) \propto P(D|AG) \times P(AG)$

$\propto 0.0369 \times 0.246 = 0.0090774$

I|| Mutations

PURINES – A, G

PYRIMIDINES – C, U, T

SUBSTITUTION (SNPs)

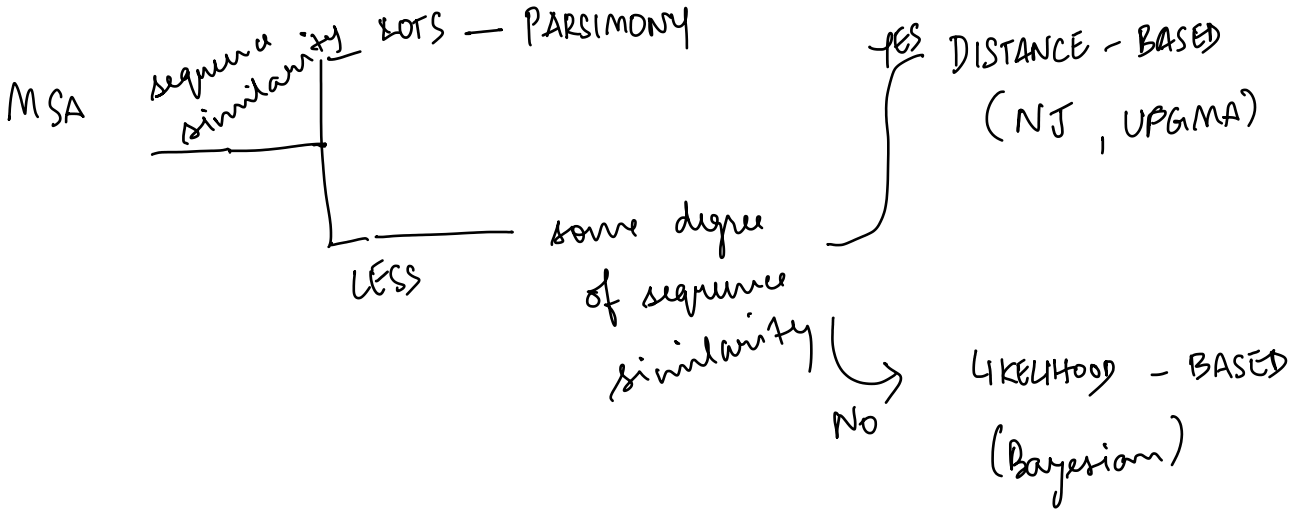
↳ transitions (PUR ↔ PUR, PYR ↔ PYR)] transitions more likely

↳ transversions (PUR ↔ PYR)]

---

synonymous vs non-synonymous] synonymous more likely

II|| INDELS] indels are less likely than subs.



PARSIMONY – MINIMUM EVOLUTION

1 | 2 | 3 | 4

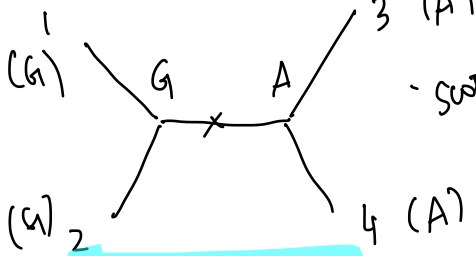
SITE 3 – invariant

INDIV				
1	A	G	T	A
2	C	G	T	G
3	T	A	T	G
4	G	A	T	G

SITES 1, 2, 4

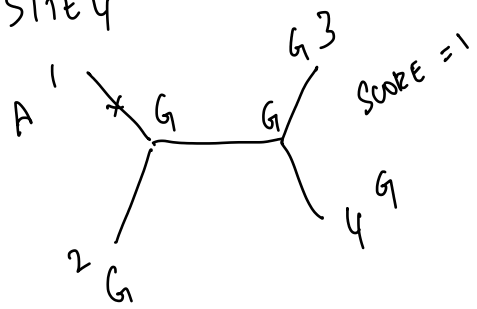
- SEGREGATING

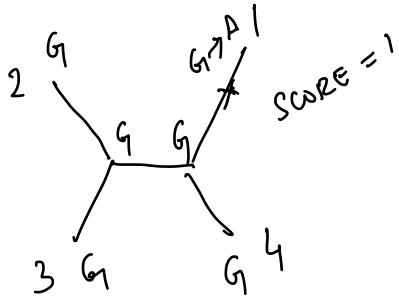
SITE 2

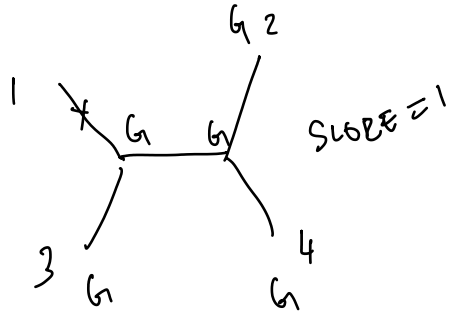


SCORE = 1

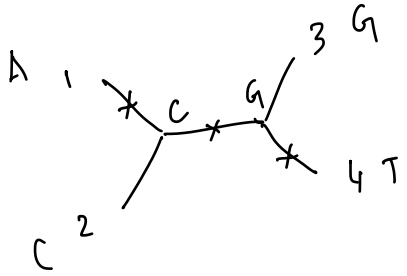
SITE 4





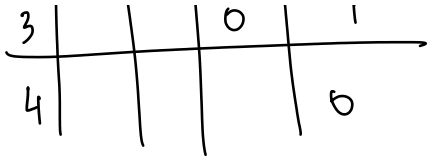


SITE 1



	1	2	3	4
1	0	2	3	3
2		0	2	2

FITCH & MARGOLIASH


3
0
1
4
6