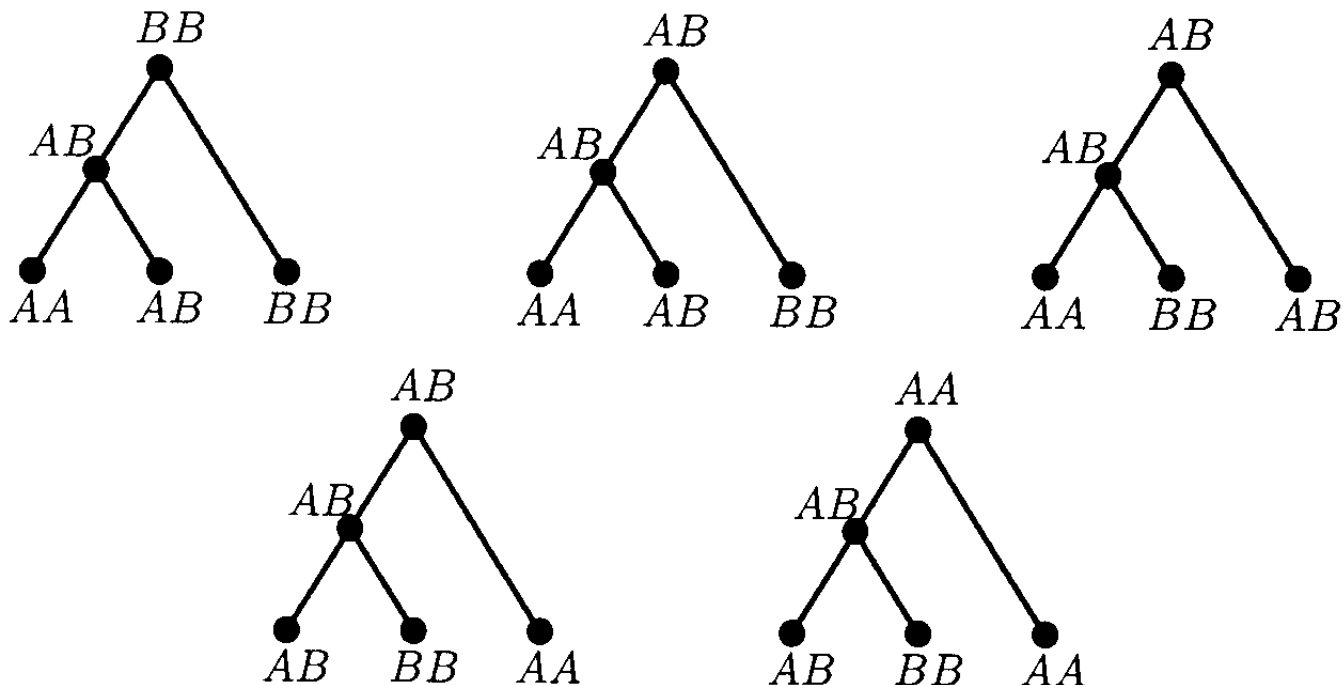


Accepted Point Mutation (Dayhoff et al. 68,72,78)

- “An APM in a protein is a replacement of one AA by another accepted by evolution”
- We want to estimate the
 - probability that given a site with AA A has undergone an APM, the new AA is B
 - the rates each AA undergoes an APM
- Dayhoff et al. estimated those from hypothetically constructed phylogenetic trees
 - originally phylogenetic trees were used to represent evolutionary relationship between species
 - they can be used to represent relationship between sequences
 - trees relating the sequences in 71 families were constructed using the *parsimony* method

The parsimony method for phylogenetic trees

- Look for a tree that can relate the observed sequences with a minimal number of substitutions
 - typically it is not unique
- An example of the most parsimonuous phylogenetic trees for the family of sequences AA , AB , BB :



Estimating transition probabilities from trees

- The transition frequencies were estimated from the neighboring sequences on the phylogenetic trees:
 - If A and B are aligned in two nodes of the tree connected by an edge then the $A \rightarrow B$ and the $B \rightarrow A$ counts are incremented
 - Within each of the 71 families the counts are averaged over all possible most parsimonious trees
- The 71 families considered had the property that any pair of sequences in them agreed in $\geq 85\%$ of the sequences
 - This restriction hopefully reduced to negligible the number of edges along which two APMs occurred in the same site
- Dividing those counts by the total number of times A mutated yields an estimate of the conditional probability that A mutated to B given that it mutated

Estimating the “mutability” from trees

- Dayhoff et al. estimated the rates at which an AA undergoes mutation by dividing the number of times it mutated by the number of times it appears in the phylogenetic trees
- They define the Markov chain transition matrix:

$$p_{AB} = m_A \frac{T_{AB}}{\sum_{C \neq A} T_{AC}}$$

Transition matrix for PAM1 ($\times 10^4$)

		ORIGINAL AMINO ACID										
		A	R	N	D	C	Q	E	G	H	I	L
		Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu
MUTANT AMINO ACID	A Ala	9867	2	9	10	3	8	17	21	2	6	4
	R Arg	1	9913	1	0	1	10	0	0	10	3	1
	N Asn	4	1	9822	36	0	4	6	6	21	3	1
	D Asp	6	0	42	9859	0	6	53	6	4	1	0
	C Cys	1	1	0	0	9973	0	0	0	1	1	0
	Q Gln	3	9	4	5	0	9876	27	1	23	1	3
	E Glu	10	0	7	56	0	35	9865	4	2	3	1
	G Gly	21	1	12	11	1	3	7	9935	1	0	1
	H His	1	8	18	3	1	20	1	0	9912	0	1
	I Ile	2	2	3	1	2	1	2	0	0	9872	9
	L Leu	3	1	3	0	0	6	1	1	4	22	9947

PAM X vs. % identity

Observed Percent Difference	Evolutionary Distance in PAMs
1	1
5	5
10	11
15	17
20	23
25	30
30	38
35	47
40	56
45	67
50	80
55	94
60	112
65	133
70	159
75	195
80	246
85	328

PAM 160 vs. BLOSUM 62

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
	0	-1	1	0	2	1	1	2	1	2	0	0	2	4	1	5	1	2	-2	5	C
		2	0	-2	0	-1	0	0	0	1	0	0	0	1	0	1	-1	1	1	-1	S
C	9		2	-1	-1	-1	0	0	0	0	0	0	-1	0	-1	1	0	1	1	3	T
S	-1	4		2	-2	-1	-1	0	0	-1	-1	-1	1	1	0	-1	0	0	2	1	P
T	-1	1	5		2	-1	-2	-2	-1	0	0	1	1	0	0	1	0	1	1	2	A
P	-3	-1	-1	7		2	0	-1	-2	0	1	1	0	0	-1	0	-1	1	2	4	G
A	0	1	0	-1	4		3	-1	-1	0	0	1	-1	0	-1	0	-1	0	0	0	N
G	-3	0	-2	-2	0	6		2	-1	-1	-1	0	-1	0	0	0	0	2	1	3	D
N	-3	1	0	-2	-2	0	6		1	0	0	2	2	1	-1	0	0	2	2	4	E
D	-3	0	-1	-1	-2	-1	1	6		0	-2	0	1	1	-1	0	0	1	3	3	Q
E	-4	0	-1	-1	-1	-2	0	2	5		2	-1	0	1	0	-1	0	1	2	2	H
Q	-3	0	-1	-1	-1	-2	0	0	2	5		-1	-1	0	-1	1	0	1	3	-4	R
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8		1	-2	-1	1	1	2	3	1	K
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5		-2	-1	-1	0	1	2	4	M
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5		-1	1	0	0	1	3	I
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5		-1	0	-1	1	2	L
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4		0	1	2	4	V
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4		-1	-2	1	F
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4		-1	2	Y
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6		-1	W
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7		
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11	
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	