Exercises
November 11, 2003

1. The Markov matrices that describe real DNA mutation tend to have their largest entries along the main diagonal in the (1,1), (2,2), (3,3), and (4,4) positions. Why should this be the case?

2. An ancestral DNA sequence of 40 bases was

$$CTAGGCTTACGATTACGAGGATCCAAATGGCACCAATGCT,$$

but in a descendent it had mutated to

$$CTACGCTTACGACAACGAGGATCCGAATGGCACCATTGCT.$$

a. Compute the JC distance between the sequences.

b. Give an initial base distribution vector and a Markov matrix to describe the mutation process.

c. These sequences were actually produced by a Jukes-Cantor simulation. Is that surprising? Explain.

3. Data from two comparisons of 400-base ancestral and descendent sequences are shown below in tables.

$S_1 \diagdown S_0$	A	G	C	T
A	92	15	2	2
G	13	84	4	4
C	0	1	77	16
T	4	2	14	70

$S_1' \diagdown S_0'$	A	G	C	T
A	90	3	3	2
G	3	79	8	2
C	2	4	96	5
T	5	1	3	94

For one of these pairs of sequences a Jukes-Cantor model is appropriate. Which one, and why?

4. Suppose we wish to model molecular evolution not at the level of DNA sequences, but rather at the level of the proteins that genes encode.

a. Create a simple one-parameter mathematical model (similar to the Jukes-Cantor model) describing the process. You will need to use that there are 20 different amino acids from which proteins are constructed in linear chains.

5. Suppose you have compared two sequences S_α and S_β of length 1000 sites and obtained the data in the table below for the number of sites with each pair of bases.

$S_\beta \diagdown S_\alpha$	A	G	C	T
A	105	25	35	25
G	15	175	35	25
C	15	25	245	25
T	15	25	35	175

a. Assuming S_α is the ancestral sequence, find an initial base distribution $\mathbf{p}_0$ and a Markov matrix M to describe the data. Is your matrix M Jukes-Cantor? Is $\mathbf{p}_0$ an equilibrium distribution for M?

b. Assuming S_β is the ancestral sequence, find an initial base distribution $\mathbf{p}_0'$ and a Markov matrix M' to describe the data. Is your matrix M' Jukes-Cantor? Is $\mathbf{p}_0'$ an equilibrium distribution for M'?

You should have found that one of your matrices was Jukes-Cantor and the other was not. This can't happen if both S_α and S_β have base distribution $(.25, .25, .25, .25)$.

6. Show the product of two Jukes-Cantor matrices is again a Jukes-Cantor matrix as follows: Let $M(\alpha_1)$ be the Jukes-Cantor matrix with parameter α_1, and $M(\alpha_2)$ the Jukes-Cantor matrix with parameter α_2. Compute $M(\alpha_1)M(\alpha_2)$ to show it has the form $M(\alpha_3)$. Give a formula for α_3 in terms of α_1 and α_2.