Genome evolution – Ex 1 Due: March 11

Solve 6 questions for 120 points. Work independently, or in pairs. You can use any available resource (books, papers, the web), but please cite what you use.

1. Compute the exponential of the matrix: $\begin{pmatrix} \alpha & \gamma \\ 0 & \beta \end{pmatrix}$

2. Describe a non-reversible Markov process in a natural or engineered environment of your choice. Formalize the process using a Q matrix and show explicitly it is non-reversible.

3. After N rolls of a dice, we obtained $n_1, n_2, .., n_6$ outcomes for the 6 dice's sides. Show that the maximum likelihood estimator for the probabilities of the dice are $\frac{n_i}{N}$.

4. A population of N bacteria is evolving under the influence of new antibiotics. The bacteria genome is of size L. At each generation, each bacterium is divided into two, mutating m% of its genome. At each generation, a fraction $e>0.5$ of the bacteria are killed by the antibiotics. If there is one specific mutation that is needed in order to achieve resistance, what is the probability that the population will acquire it? What if two mutations are needed? Three?

Analytic, simulation or approximate solutions are acceptable.

5. You are trying to learn a tree model for a triplet phylogeny as shown in class (h1,h2,s1,s2,s3), using 4 different **probability** (not rate) matrices for 4 lineages (h2->h1,h2->s3,h1->s1,h1->s2). You assume that the probability matrices are symmetric. Is it possible to learn uniquely two different matrices for the two branches leaving the root? Prove (you are asked to determine if the observations on the S's provide enough constraint to resolve the different substitution patterns at the two root lineages, if the matrices are symmetric. Prove it can be done, or give a counter-example where two solutions give the maximum likelihood)

6. You are again trying to learn a tree model for a triplet phylogeny as shown in class (h1,h2,s1,s2,s3), using 4 different matrices for 4 lineages (h2->h1,h2->s3,h1->s1,h1->s2). You do not assume symmetry or reversibility, but you do assume that the substitution probabilities in h2->s3 are twice as much as those in h2->h1 (Pr(s3=y|h2=x)=2*Pr(h1=y|h2=x), for x!=y). Can this be solved uniquely? Write down the EM maximization formula if the answer is yes, give a counter example if the answer is no.

Good luck!