Rubric and answers for BC homework 1

Question 1: (total = 2)

A statement you are aware of the honor code policies from CSE and that collaboration is fine, but copying answers from others or online resources is not.  Please note in your report or comments whomever you may have discussed the topic with outside of what we will cover in class to be safe.  
  
Question 2:   
  
1 point : lambda sequence with accession NC\_001416 in dropbox named as lambda.fasta  
  
Question 3: (total = 5)  
  
1 point : outputs a new header of ">reversed"  
2 points : correctly reverses sequence  
2 points : correctly complements sequence  
  
Question 4: (total = 5)  
  
1 point: correctly reads in FASTA sequence  
2 points: correctly computes nucleotide freqs (see below)  
2 points: correctly computes dinucleotide freqs (see below)  
  
Question 5:   
  
1 point : human mitochondria sequence with accession NC\_012920 in dropbox named as human\_mito.fasta  
  
Question 6: (total = 8)  
  
3 points: correctly computes probability using multinomial  
2 points: correctly builds markov model of order 3  
3 points: correctly computes probability using markov model   
  
Question 7: (total = 8)  
  
3 points: correctly builds markov model (same as #5)  
1 points: correctly starts sequence generation (first nucleotide)  
2 points: correctly continues sequence generation (next 19,999 nucleotides)  
2 points: generates a sequence of 20,000 preferably in FASTA format

**As for answers replacing Ns with As:**

Answer to problem 3 (no need to include these many decimals... it is what Perl does by default):  
-------------------------  
  
Nucleotide frequencies:  
A 0.2542987918024  
C 0.234258381097687  
G 0.264318997154756  
T 0.247123829945157  
  
Dinucleotide frequencies:  
AA 0.076122141811509  
AC 0.0530504525679883  
AG 0.0563287354899899  
AT 0.0688027050988639  
CA 0.0663079111770891  
CC 0.0514834745675347  
CG 0.0641842436238428  
CT 0.0522875816993464  
GA 0.0671326364404858  
GC 0.0745345456794705  
GG 0.0655656584400322  
GT 0.0570709882270469  
TA 0.0447413455392672  
TC 0.0551947382528195  
TG 0.0782251912331705  
TT 0.0689676501515433

Perl based answers for probabilities are:  
  
Log probability (multinomial): -452.662312232692  
Log probability (Markov): -451.796605782161