

Molecular Evolution & Phylogenetics (BIOS 659: Advanced Studies in Genetics)

Exercise II – 9 Sept 2004

You'll need to use MEGA3 for the following problems. Mega is installed on the PC's – you'll find it under start/programs/Mega3.

1. The amino acid sequences of hemoglobin α chains from the human, horse, cow, kangaroo, newt, and carp are available from the MEP book website: <http://lsvl.la.asu.edu/mep/>.

- a) Eliminating all alignment gaps, compute the p-distance, Poisson correction (PC) distance, and the gamma distance with $\alpha = 2$ and $\alpha = 1.3$ for all pairwise comparisons of sequences. Use the MEGA3 program and give four digits after the decimal point.
- b) Compute the standard errors of the above four distances using both analytical and bootstrap methods.

2. The nucleotide sequence of the Cytochrome *b* (cytb) gene from human and Rhesus monkey are available from the MEP book website: <http://lsvl.la.asu.edu/mep/>.

- a) Count the number of differences between the sequences and calculate the proportion of differences (p).
- b) Estimate the number of substitutions between the sequences by “correcting” p with the following models of nucleotide substitution:
 1. Jukes-Cantor (JC)
 2. Kimura's 2-parameter (K2P)
 3. Felsenstein (F81; aka Tajima-Nei or “equal input”)
 4. Hasegawa, Kishino & Yano (HKY; aka Tamura-Nei)
 5. General reversible model (REV)

Compare the numbers of substitutions estimated by the various models to the observed number.

- c) Estimate the number of substitutions between the sequences by “correcting” p with the following models of nucleotide substitution with gamma distributions for substitution rate variation:
 1. JC+ Γ
 2. K2P+ Γ
 3. F81+ Γ
 4. HKY+ Γ

For the gamma shape parameter, use $\alpha=0.5$ (wide variation in rates, typical for this gene) and $\alpha=2$ (bell-shaped distribution of rates). Compare the numbers of substitutions estimated by the various models to the observed number and note the affect of the shape parameter.

- d) For the JC and the HKY models, calculate the standard error using the analytical method and the bootstrap methods with 100 and 1000 replicates. How does the estimated standard error vary with the three measures?
- e) Compare estimates of the K2P and HKY models (with standard errors) at the 1st, 2nd, and 3rd positions. How do the estimates of the numbers of substitutions compare to each other and to the observed number of estimates? How can these be interpreted given the standard errors for the two corrections?