Bioinformatics	(Akiyama	) Exercise #05

Student ID	Name
------------	------

 Calculate E-value in the following case of ungapped local alignment. Query protein sequence was length of 400 a.a., and target DB was nr-aa database which has total sequence length of about 2.5 × 10<sup>9</sup>. A hit sequence with Bit score = 30 (bits) was obtained. Then what is the E-value (expected hit counts with score >= 30 bits) ? Based on Karlin-Altschul theory, E = m n 2<sup>-S'</sup> (m,n : length, S': bit score). Here, an apploximation of <u>2<sup>10</sup> ≒ 10<sup>3</sup></u> must be used, for simplicity.

## Ans.

2) Calculate E-value in the case of ungapped local alignment. Query length and database size are equal to exercise 1). What is the E-value for Bit score =50(bits)? Use the apploximation of  $2^{10} \approx 10^3$ .

Ans.

3) Calculate P-value in the following case of ungapped local alignment.
P-value can be calculated from an E-value by the equation: P = 1 - exp (-E).
Ungapped local alignment with a query sequence length = 100, and
database total length = 10,000, gave a hit with Bit score = 19 (bits).

Then what is the P-value (probability of obtaining a hit with score >= 19 bits)? Here, an apploximation of  $2^{10} \approx 10^3$ , and  $\underline{e} \approx 3$ , must be used for simplicity. Answer with one significant figure is enough.

Ans.

4) Calculate P-value. Conditions are equal to exercise 3).
What is P-value for Bit score = 50 (bits).
Here, use the apploximation of 2<sup>10</sup> ≒ 10<sup>3</sup>.
Answer with one significant figure is enough.