

問題番号 [1 8] <情報生物学>

ゲノムの持つ情報量に関する以下の文章について、下の設問 (1) ~ (6) に答えなさい。(計 50 点)

ゲノムの持つ情報量について考える。情報量は一般にビットという単位で表される。1 ビットは 0 か 1 の 2 状態のどちらかであるかが判明したときに得られる情報として定義される。ゲノム配列中のある部位における情報量 H は次に示す式 1 によって計算される。

$$H = - \sum_b p_b \log_2 p_b \quad \text{式 1}$$

ここで、 p_b は塩基 b の出現割合を表す。4 種類の塩基の出現割合が等しいと仮定すると、ゲノム配列中の各部位はそれぞれ (A) ビットの情報量を持っていると言える。

ヒトのゲノムサイズはおよそ (B) 塩基であることが知られている。ここで、先ほどの仮定、すなわち(a) 4種類の塩基の出現割合が等しいとし、また 8 ビット = 1 バイトであるためバイトに換算すると、全ゲノムが有する情報量は (C) メガバイトとなる (メガは 10^6 であることを示す)。ただし、タンパク質として翻訳されるような(b) 機能的に重要な領域は、全ゲノム配列中の (D) %程度と言われているため、実際に有効な情報量は (E) メガバイト程度となる。

(1) 文中の (A) に入る数字を答えなさい。(10 点)

(2) 文中の (B) に入る数字としてもっとも適当なものを次の中から選び、記号で答えなさい。(5 点)

(ア) 3×10^6 、(イ) 3×10^7 、(ウ) 3×10^8 、(エ) 3×10^9

(3) 文中の (C) に入る数字を答えなさい。(5 点)

(4) 文中の (D) に入る数字としてもっとも適当なものを次の中から選び、記号で答えなさい。(5 点)

(ア) 2、(イ) 30、(ウ) 50、(エ) 98

(5) 文中の (E) に入る数字を答えなさい。(5 点)

(6) 下線(b)の領域において下線(a)が成り立たない場合、(E) の値がどうなるか、例を挙げつつ述べなさい。なお、必要に応じて次の値、 $\log_2 0.2 = -2.32$ 、 $\log_2 0.3 = -1.74$ 、を用いること。(20 点)