

情報生物学 (1 / 4)

(注意) 問題 [1] [2] [3] [4] はそれぞれ別の答案用紙に解答すること。

裏も使用してよい。解答は英語、日本語どちらでも可とする。

[1] 以下の遺伝子変異に関する問題について答えなさい。(計 25 点)

ヒトの体細胞では、DNA の二本鎖切断の非相同末端結合による修復により、1 年間で細胞 1 個あたりおよそ 3 つの変異が蓄積しているとの報告がある。一方で、実験室では PCR などを用いて人為的に遺伝子に変異を導入する場合はある。

(1) ヒトゲノムのうち 1.5% がタンパク質をコードしている場合、日本人の平均寿命である 80 歳ではタンパク質をコードする遺伝子領域に細胞あたり何個程度の変異が挿入されると考えられるか? 小数点第一位まで答えなさい。(10 点)

(2) PCR を用いて 1% の確率で特定の遺伝子に変異を導入したところ、あるチロシン (TAT) の 3 番目の塩基に変異が導入され終止コドンに変化していた。なお、チロシン (TAT) の三番目に変異が導入された場合は、チロシンか終止コドンのいずれかしか選ばれない。

(a) このような変異を何と呼ぶか答えなさい (5 点)

(b) このコドンの 3 番目に変異が導入されることで終止コドンに変化する確率 (%) を小数点第二位まで答えなさい。(10 点)

情報生物学 (2 / 4)

(注意) 問題 [1] [2] [3] [4] はそれぞれ別の答案用紙に解答すること。
裏も使用してよい。解答は英語、日本語どちらでも可とする。

[2] 次の文章を読んで以下の問いに答えなさい。(計 25 点)

次に示すのは、ヒトとマウスのある遺伝子の 20 塩基長の部分塩基配列とその共通祖先の配列とのアラインメントである。3 つの配列で塩基の異なる座位を四角で囲ってある。

ヒト	TGCGGCTC	A	CACCT	G	GTGGA
マウス	TGCGGCTC	C	CACCT	A	GTGGA
共通祖先	TGCGGCTC	G	CACCT	G	GTGGA

ヒトとマウスだけを比べると、それらの間では 2 つの座位が異なるので、全座位あたりの置換数は 0.1 であるように見える。しかし共通祖先も含めて考えると、共通祖先からの分岐後に 3 回の置換が起きたと考えるのが妥当である。なぜなら、左から 9 番目の座位ではヒトとマウスで独立に置換が起きているからで、このような置換を (A) と呼ぶ。2 本の配列の比較から正確な置換割合を求める上で、それらの配列が (B) ほど、この (A) の影響が無視できなくなる。下に挙げる式 1 を用いることで、配列間の見かけの相違度 p から、実際の座位当たりの置換数 K を見積もることが可能である。

$$K = -\frac{3}{4} \ln \left(1 - \frac{4}{3} p \right) \quad \text{式 1}$$

- (1) (A) に当てはまる言葉を次の中から選び、記号で答えなさい。(5 点)
(ア) 同義置換、(イ) 非同義置換、(ウ) 復帰突然変異、(エ) 多重置換
- (2) (B) に当てはまる言葉を次の中から選び、記号で答えなさい。(5 点)
(ア) 長い、(イ) 短い、(ウ) 進化的距離に遠い、(エ) 進化的距離に近い
- (3) 2 本の DNA 配列の見かけの相違度 p が 0.25 であったとき、座位当たりの置換数 K を式 1 から見積もりなさい。ただし、挿入と欠失は起きなかったものとする。なお、必要に応じて次の値、 $\ln \frac{2}{3} = -0.4$ 、を用いること。(5 点)
- (4) 進化的に長い時間が経ち、座位当たりの置換数 K が十分大きくなると、見かけの相違度 p はどのような値に収束するか求めなさい。(10 点)

情報生物学 (3 / 4)

(注意) 問題 [1] [2] [3] [4] はそれぞれ別の答案用紙に解答すること。

裏も使用してよい。解答は英語、日本語どちらでも可とする。

[3] バイオインフォマティクスにおける遺伝子やタンパク質の配列解析について以下の問いに答えなさい。(計 25 点)

(1) アラインメントとは、複数の配列の共通領域または類似領域を特定できるように配列の文字列を並べたものである。以下のような二本のアミノ酸配列のアラインメントが得られたとする。このとき、アミノ酸残基の一致の割合で、配列類似性のスコアを計算しなさい。(5 点)

配列 1 : MNSFSTSAFP

配列 2 : MNTFSSAFGP

(2) アミノ酸間の物理化学的性質の類似性の指標として、右図のような BLOSUM62 アミノ酸置換行列と呼ばれる対応表がある。これに基づいて、上記の二本のアミノ酸配列のアラインメントにおける配列類似性のスコアを計算しなさい。(10 点)

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-2	-1	1	0	-3	-2	0	
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

(3) ギャップ記号“-”を挿入し、文字列をずらすとアラインメントを改善できることがある。上記の二本のアミノ酸配列それぞれにギャップ記号を一つだけ挿入して、より多くのアミノ酸が一致するようなアラインメントを作成しなさい。更に、そのときの配列類似性のスコアを BLOSUM62 アミノ酸置換行列に基づいて計算しなさい。ただし、アミノ酸とギャップ記号の対応 (ギャップペナルティ) は -2 とする。(10 点)

情報生物学 (4 / 4)

(注意) 問題 [1] [2] [3] [4] はそれぞれ別の答案用紙に解答すること。

裏も使用してよい。解答は英語、日本語どちらでも可とする。

[4] 以下の問いに答えなさい。(計 25 点)

特定の条件において発現が変動する遺伝子群が、ERK 経路に対して関係があるかどうかを解析したい。 N 個の遺伝子を調べた結果、特定の条件において有意に変動している遺伝子は d 個であった。 N 個の遺伝子のうち、ERK 経路に属している遺伝子は n 個であった。 n 個の遺伝子のうち、特定の条件において有意に変動している遺伝子は、 x 個であった (表)。また、 ${}_a C_b = \frac{a!}{(a-b)!b!}$,

$\binom{a}{b} = \frac{a!}{(a-b)!b!}$ のいずれの表記を用いてもよい。

表	ERK 経路に属する	ERK 経路に属さない	計
変動あり	x	$d-x$	d
変動なし	$n-x$	$N+x-d-n$	$N-d$
計	n	$N-n$	N

- (1) N 個の遺伝子から無作為に d 個の遺伝子を非復元抽出する組み合わせの数を、 N, d, n, x を用いて表しなさい。(6 点)
- (2) N 個の遺伝子から無作為に d 個の遺伝子を非復元抽出したときに、 x 個の遺伝子が ERK 経路に属して、かつ、 $d-x$ 個の遺伝子が ERK 経路に属さない組み合わせの数を、 N, d, n, x を用いて表しなさい。(6 点)
- (3) N 個の遺伝子から無作為に d 個の遺伝子を非復元抽出したときに、表のような結果が得られる確率を、 N, d, n, x を用いて表しなさい。(6 点)
- (4) 片側フィッシャーの正確確率検定 (超幾何検定) では、帰無仮説 H_0 : 「特定の条件における遺伝子群の変動と ERK 経路が独立である (関係がない)」、対立仮説 H_1 : 「特定の条件における遺伝子群の変動と ERK 経路が独立でない (関係がある)」として、検定を行う。 p 値を、 N, d, n, x を用いて表しなさい。ただし、 $d > n$ とする。(7 点)