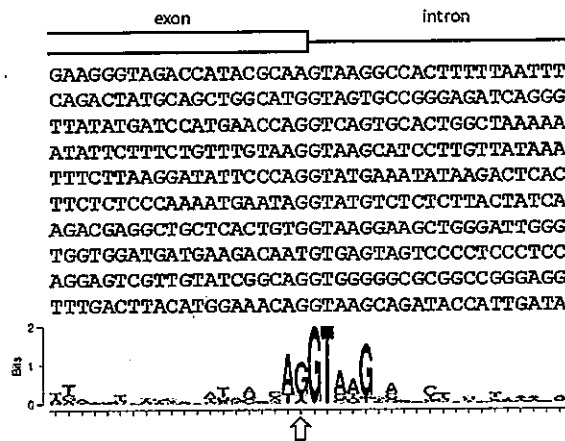


情報生物学 (1 / 2)

(注意) 問題 [1] [2] はそれぞれ別の答案用紙に解答すること。

- [1] 以下に示すのは、ヒトゲノムから任意に選んだ 10ヶ所のエクソン-イントロン境界のゲノム塩基配列のアラインメントである。アラインメントの下には各サイトでの塩基の現れやすさがシーケンスロゴの形で表示されている。このアラインメントに関する以下の問いに答えなさい。(計 50 点)



- (1) 5'スプライス部位には特徴的な 2 塩基配列 GT が見られる。なぜそうなるのか、また、スプライシングにおいてこの 2 塩基パターン以外にも重要と考えられる配列上の特徴についてこの図からわかることを述べなさい。(15 点)
- (2) この 2 塩基配列 GT に変異が起きた場合、発現した転写産物がどうなると考えられるかについて述べなさい。(15 点)
- (3) シーケンスロゴは各サイトでの塩基の現れやすさを情報量を用いて表したものである。アラインメント中 i 番目のサイトが有する情報量 C_i は次に示す式 1 によって計算される。情報量の単位はビットである。

$$C_i = 2 + \sum_b p_{b,i} \log_2 p_{b,i} \quad \text{式 1}$$

ここで、 $p_{b,i}$ は、 i 番目のサイトにおける塩基 b の割合を表す。ただし、ここで $p_{b,i} = 0$ のとき $p_{b,i} \log_2 p_{b,i} = 0$ とみなす。上の図と式 1 からわかるように、あるサイトが 1 種類の塩基のみで占められている場合、そのサイトの情報量は 2 ビットとなる。また、 i 番目のサイトで塩基 b の文字を図示する際の高さ $H_{b,i}$ は次に示す式 2 によって計算される。

$$H_{b,i} = C_i p_{b,i} \quad \text{式 2}$$

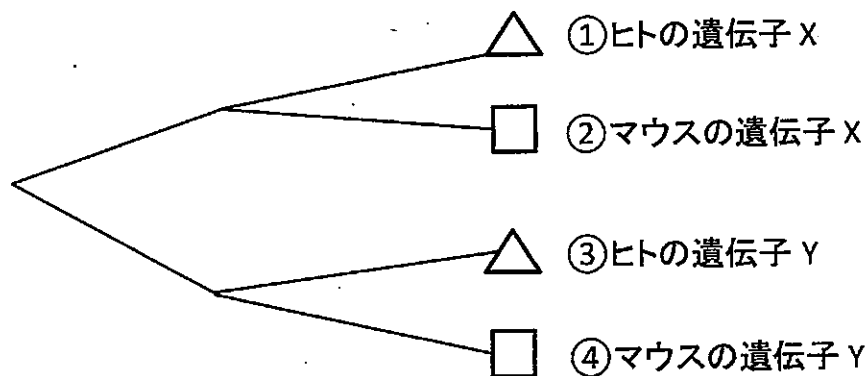
式 1 と式 2 を使い、矢印で示したエクソンの 3' 端のサイト (塩基の頻度は、A = 1, C = 0, G = 8, T = 1) の情報量と各塩基の文字を図示する際のそれぞれの高さを求めなさい。なお、必要に応じて次の値、 $\log_2 0.1 = -3.32$ 、 $\log_2 0.8 = -0.32$ 、を用いること。(20 点)

情報生物学 (2 / 2)

(注意) 問題 [1] [2] はそれぞれ別の答案用紙に解答すること。

[2] 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 50 点)

- (1) 配列の相同性の観点からオーソログとは何か説明しなさい。(10 点)
- (2) 配列の相同性の観点からパラログとは何か説明しなさい。(10 点)
- (3) 遺伝子①、②、③、④に関する以下の様な系統樹がある場合 (ヒトの遺伝子は三角、マウスの遺伝子は四角で表す)、オーソログの関係にある遺伝子はどれか、パラログの関係にある遺伝子はどれかそれぞれ答えなさい。(10 点)



- (4) 配列の類似度検索で行う配列アライメントと機能解析や進化解析との関連について説明しなさい。(10 点)
- (5) 配列の類似度検索を用いて、計算機的にオーソログを同定するにはどのような方法が考えられるか説明しなさい。(10 点)