

平成26年度

日本バイオインフォマティクス学会 (JSBi)

バイオインフォマティクス技術者認定試験

解説集

問 1 正解【3】

LINE (Long interspersed nuclear element) は長鎖散在反復配列である。染色体そのものではなく、ゲノム中の遺伝子が整数倍化することを遺伝子重複と呼ぶ。マルチジーンファミリーは遺伝子重複によって生じた遺伝子ファミリーを意味する。

問 2 正解【3】

質的形質では、表現型が不連続で明確な区分が可能であり遺伝子座との関係が明らかにされやすい。一方、量的形質では、表現型が連続した値をとり、複数の遺伝子座と関わっている。QTL解析は、個体の量的形質を支配する遺伝子を探索する方法である。

問 3 正解【4】

オートファジーは、細胞内の不要タンパク質等をリソソームに輸送して分解する機構である。原核細胞はオートファジーの機構を持たない。

問 4 正解【1】

真核生物のゲノム DNA は環状ではなく直鎖状で存在し、ヒストンタンパク質に巻き付いて染色体を構成している。

問 5 正解【4】

DNA の五炭糖はデオキシリボース、RNA の五炭糖はリボースである。

問 6 正解【4】

能動輸送は、ATP を利用した膜内外の物質の濃度勾配に逆らう輸送である。

問 7 正解【1】

膜貫通領域に  $\alpha$  ヘリックス構造をもつ膜タンパク質には、例えばチャネル、トランスポーター、受容体などがある。一方、膜貫通領域が  $\beta$  シート構造で折り返された  $\beta$  バレル構造をとるものには、バクテリアのポーリンやホスホリパーゼ A などがある。

問 8 正解【4】

液胞は植物細胞にみられる細胞内小器官であり、1 枚の脂質二重膜からなる。細胞内の代謝産物を貯蔵する機能を持つ。

問 9 正解【1】

細胞周期は、DNA 合成の準備期 (G1 期)、DNA 合成期 (S 期)、細胞分裂の準備期 (G2 期)、細胞分裂期 (M 期) の順で繰り返される。

問 10 正解【4】

減数第一分裂では、染色分体の間で乗換え（交差）が生じ、父方由来と母方由来の DNA 配列が部分的に交換される。この結果、同一染色体にある遺伝子の組み合わせが変化し、子孫に遺伝的多様性を生むことになる。

問 11 正解【4】

DNA ポリメラーゼによる DNA 鎖の伸長反応にはプライマーが必要である。

問 12 正解【4】

DNA にある遺伝情報が mRNA に転写され、リボソーム内でタンパク質に翻訳されることで遺伝情報の発現が行われる。この流れの概念をセントラルドグマという。

問 13 正解【1】

パルミトイル化はシステイン残基への脂質修飾、O-結合型グリコシル化はセリン・スレオニン残基への糖修飾である。リン酸化はセリン・スレオニン・チロシン・ヒスチジンへのリン酸基の付加、ジスルフィド化はシステイン残基間の架橋形成である。

問 14 正解【4】

負電荷アミノ酸に分類されるのはアスパラギン酸・グルタミン酸、正電荷アミノ酸はアルギニン・リジン・ヒスチジンである。ヒスチジンは、pH 5 の水溶液中では側鎖からプロトンが奪われ、極性非電荷アミノ酸となる。最も分子量の大きいアミノ酸は、芳香環を持つトリプトファンである。

問 15 正解【3】

タンパク質からなる酵素は、pH、塩濃度、溶媒などの条件の変化によって機能活性が大きく異なる。

問 16 正解【3】

クエン酸回路および酸化リン酸化は、ともにミトコンドリア内の反応である。

問 17 正解【3】

神経型では、神経細胞から放出されたシグナル分子が、シナプス部で接触している標的細胞の受容体に結合することにより、シグナルが伝わる。

問 18 正解【3】

グリコシド結合とは、糖のヒドロキシル基と他の糖やアルコールのヒドロキシル基とが脱水縮合して形成する共有結合である。タンパク質における $\pi$ - $\pi$ 相互作用は、芳香環をもつアミノ酸（ヒスチジン、フェニルアラニン、チロシン、トリプトファン）同士が近接して生じる非共有結合である。

問 19 正解【2】

体液性免疫においては、ヘルパーT細胞から抗原提示を受けたB細胞が抗体を産生する。

問 20 正解【3】

10 サイクルのPCRで、1分子の2本鎖DNAが $2^{10}$ 倍になる。

問 21 正解【2】

nの補数を取ると $-n$ になる。nと $-n$ は等しくない。

問 22 正解【2】

CPUの周波数を2倍にしても、メモリに対するアクセス速度が遅い場合などは、必ずしも2倍の速度が出ない。

問 23 正解【1】

ネットワークに接続し、インターネットを閲覧する際には、メールアドレスは必要ない。

問 24 正解【4】

FTP, TELNET, SMTPはいずれも暗号化されないため、秘匿性はない。SSHは暗号化される。

問 25 正解【3】

ヒープ構造は木構造を用いて効率よく実装できる。リストではない。

問 26 正解【2】

擬似コードは深さ優先探索、かつ、到達したノードから順に表示していくプログラムになっている。

問 27 正解【3】

$O(n)$ 時間アルゴリズムを $O(n)$ 回実行したら、 $O(n^2)$ 時間を要する。

問 28 正解【3】

C, D を先に取り出せるように操作した場合、最後の 2 つは B, A の順になってしまう。

問 29 正解【2】

2 分探索では、n 回で最大  $2^{(n+1)} - 1$  個の値を発見することが可能である。

問 30 正解【2】

平均リード長が 100 を超え、エラー率が 1.0 を下回る実験番号は、HX331 と RE071 の 2 件

問 31 正解【1】

XML は各要素をタグで囲って表す。一般に表では表さない。

問 32 正解【3】

リレーショナルデータベースでは、複数の表に格納する。

問 33 正解【4】

1 は X と Y が独立の時のみ成り立つので、常には成り立たない。2 と 3 は、相関と期待値は関係がないので、いずれも不適切。

期待値演算子を E で、X と Y の共分散を  $\text{Cov}[X, Y]$  で表すとする。定義より

$$\text{Var}[X] = E[X^2] - (E[X])^2$$

$$\text{Var}[Y] = E[Y^2] - (E[Y])^2$$

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

が成り立つ。選択肢 4 は次のように証明できる。

$$\begin{aligned} \text{Var}[X + Y] &= E[(X + Y)^2] - (E[X + Y])^2 = E[X^2 + Y^2 + 2XY] - (E[X] + E[Y])^2 = E[X^2] + \\ &E[Y^2] + 2E[XY] - (E[X])^2 - (E[Y])^2 - 2E[X]E[Y] = \text{Var}[X] + \text{Var}[Y] + \text{Cov}[X, Y] > \text{Var}[X] + \\ &\text{Var}[Y] \end{aligned}$$

問 34 正解【2】

空白(a),(b)の解説：中心極限定理より次のことが言える。母平均  $\mu$  および母分散  $\sigma^2$  を持ついかなる連続確率分布  $p$  であっても、 $p$  から独立試行によって得た標本観測値の標本平均  $\bar{x}$  は、標本数を増やすほどその分布は正規分布に漸近する。標本数を  $n$  とすると、 $\bar{x}$  の平均は  $\mu$  となり、 $\bar{x}$  の分散は  $\sigma^2/n$  となる。

空白(c)の解説：正規分布に従う確率変数の線形結合は正規分布に従う。標本平均  $\bar{x}$  は  $n$  個の確率変数の線形結合なので、 $n$  の値に依らず正規分布に従う。

問 35 正解【4】

空白(a)の解説：平均がある値  $\mu_0$  に等しいか検査する場合、「両側検定」を行い、平均がある値  $\mu_0$  より大きい検査する場合、もしくは小さい検査する場合、「片側検定」を行う。  
空白(b),(c)の解説：仮説検定において P 値が有意水準を下回れば帰無仮説は棄却される。

問 36 正解【4】

選択肢 4 は、正しくは  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  である。

ちなみに選択肢 3 の  $P(A^c) = 1 - P(A)$  は明らかに正しい。

ドモルガン法則  $(A \cap B)^c = A^c \cup B^c$ ,  $(A \cup B)^c = A^c \cap B^c$  それぞれと選択肢 3 を組み合わせると、選択肢 1 と選択肢 2 が導かれる。

問 37 正解【2】

多くの確率分布では、確率密度関数の値は 1 を超える場合がある。スチューデント t 分布の場合、分散が小さければ確率密度関数の値は 1 を超えることがある。一方、累積密度関数では、その定義から、いかなる確率分布でも 1 を超えることはない。

問 38 正解【1】

精度と再現率は、それぞれ  $TP/(TP + FP)$  および  $TP/(TP + FN)$  で定義される。F 値は精度と再現率の調和平均で与えられる。精度の定義にも再現率の定義にも TN が含まれないことから、精度・再現率・F 値の算出に TN は不要である。

問 39 正解【3】

分割統治法はアルゴリズムの戦略の一つで、そのままでは解けないような大きな問題を小さな問題に分割して各個解いていく方法論を指す。

問 40 正解【2】

入力層、中間層、出力層の 3 層。

問 41 正解【2】

大量 cDNA 解析の品質検査には、配列データのクオリティをチェックする 1 のほかに、発現が期待されるマーカー遺伝子が検出されているかを確認したり、他生物種由来の配列の混入がないかを確認したりすることが含まれる。2 のような通常ホモロジー検索は、配列データ解析としては定番の手順だが、品質検査としての目的が不明確であるため、不適。

問 42 正解【3】

バイサルファイトシーケンスでは、メチル化されていないシトシンのみがチミンに置換されるため、シトシンとしてシーケンスされた部位はメチル化されたシトシンであったと推定できる。近年、次世代シーケンスを用いることによって、全ゲノム規模でのメチル化解析（メチローム解析）が進展し、選択肢 1, 2, 4 の内容を含めて様々な事実が明らかになりつつある。選択肢 3 はメチル化と置換の有無の関係が逆なので間違い。

問 43 正解【2】

エキソームシーケンスは、全ゲノム DNA 由来のサンプルから、既知のエキシソンの配列情報を使ってエキソン領域を含む配列断片のみを濃縮し、これをシーケンスすることによって、エキソン領域における変異情報を効率的に収集する方法である。既知のエキシソンが対象となるため、未知のエキシソンの位置を調べる目的には使えない。

問 44 正解【2】

COGs は、ゲノム間で遺伝子の総当たりのホモロジー検索を行って、双方向ベストヒットをとるなどの基準を用いてオーソログ関係を推定し、構築されているが、最終的には手作業を交えた分類が行われている。

問 45 正解【4】

メタゲノム解析では、ホモロジー検索によって遺伝子機能と由来生物種の推定を行うが、一般に由来生物種を推定するには高い類似性でのヒットが必要となる。一方、原核生物ではゲノムごとに特徴的な塩基配列組成を持っているため、塩基配列組成に基づいて同一生物種由来と推定される配列をまとめることが可能である。その際は適当な長さの固定長塩基配列頻度を用いるのが普通だが、簡易的な指標としては GC 含量を用いることもできる。GC-skew は多くの原核生物ゲノムにおいて複製方向と関連しており、複製開始・終結点の推定に有効な指標であるが、生物種ごとのゲノム配列の特徴づけには向かない。

問 46 正解【4】

水平伝搬した遺伝子は、塩基組成（GC 含量、コドン使用頻度など）や系統関係（ホモロジー検索の上位ヒットなど）が伝搬元の生物種の特徴を反映しているために、他の遺伝子とは異なり、これを水平伝搬遺伝子の検出に利用することができる。選択肢 4 は全生物の共通祖先から保存されている普遍的な遺伝子の特徴であり、水平伝搬の証拠ではない。

問 47 正解【4】

接尾辞配列は接尾辞を辞書順にソートしたものであるから、表の接尾辞のカラムに注目して、空欄となっている2つの接尾辞がそれぞれ何かを考える。可能性があるのは位置4のGATGAと位置6のTGAだが、辞書順ということから6行目がGATGA、最後の行がTGAとわかる。

問 48 正解【3】

選択肢1と2は正しく、選択肢3はこれらの関係を使って確認できる。実際、長さ5000の配列には $5000 - 5 + 1 = 4996$ 個の5-merが含まれ、5-merの種類は $4^5 = 1024$ 通りあるから、各5-merの平均的な出現頻度は $4996 / 1024 > 1$ となる。一般に $k$ を大きくするほど $k$ -merあたりの平均出現頻度は小さくなるため、無関係な配列間で同じ $k$ -merが偶然に一致する確率が下がり、索引づけによる検索の効率を上げることができる。

問 49 正解【1】

BLOSUM行列は、配列モチーフに基づいて作成された保存配列アラインメントのデータベース(BLOCKS)を用いて、それらのアラインメント中でのアミノ酸の置換頻度を集計することによって作成されている。物理化学的性質が似たアミノ酸対に高いスコアが与えられているが、それはあくまで進化的な置換頻度を反映した結果であり、直接アミノ酸の物理化学的性質を用いて定義されたからではない。従って1が間違い。なお、BLOSUMでは置換頻度を集計する際に、 $x\%$ 以上類似した配列を同一グループにまとめて、グループ間での置換頻度をとっており、この $x$ の値を付してBLOSUM40やBLOSUM80などと呼んでいる。その結果、 $x$ が大きいほど、高い類似性を持つ(進化的に近い)配列間の比較を想定したものとなるため、4は正しい。

問 50 正解【4】

SAM形式のファイルの各行には、各リードがリファレンス配列上のどの位置にどのようにマッピングされたかについての情報が記載されている。選択肢1~3はいずれもそのような情報であるが、4はマッピングされた結果を、リファレンス配列上の位置ごとに集計して得られる結果であり、SAMファイルの各行が持つ情報とは異なる。



問 51 正解【2】

3つのステム領域における相補塩基対の位置は、それぞれ(5-8と45-48)、(13-16と21-24)、(29-32と37-40)であるから、それらの位置が正しくプロットされている選択肢を選ぶ。その際、縦軸の方が横軸より小さな位置番号に対応していて、かつ相補対では向きが反対になることに注意して、「縦軸 5-8 と横軸 48-45 とが対応する領域」等を探していけばよい。なお、選択肢 3,4 はそれぞれ別の二次構造を表しているが、選択肢 1 は 1 つの領域が 2 重に相補対を形成しており、RNA の二次構造としてはあり得ない。

問 52 正解【3】

計算式に沿って計算する。全部が同一の文字である 1, 2, 5 番目については、いずれも問題文と同じ計算式でスコアは 2 となる。また、出現しない文字については、 $0 \log 0 = 0$  よりスコアが 0 になるので無視できる。よって、3 番目については  $0.5 \log_2 \frac{0.5}{0.25} \times 2 = 0.5 \times 2 = 1$ 、

4 番目については  $0.5 \log_2 \frac{0.5}{0.25} + 0.25 \log_2 \frac{0.25}{0.25} = 0.5 + 0 = 0.5$  と計算され、全部加えるとスコアは 7.5 となる。

問 53 正解【3】

RMSD は対応する原子間距離の二乗平均の平方根であるので、 $\{(2.0^2 + 1.0^2 + 0.0^2 + 1.0^2 + 3.0^2) / 5\}^{1/2} = 1.73 \dots \text{\AA}$  (選択肢 3 にもっとも近い) が正解である。

問 54 正解【1】

二面角  $\varphi$ ,  $\psi$ ,  $\omega$  は、二次構造などタンパク質の主鎖の局所構造を特徴づける指標としてよく用いられる。その定義はやや煩雑だが、回答に必要な情報は問題文に含まれている。二面角は四つの原子の並び  $a-b-c-d$  で決定され、結合  $b-c$  の回転角である。 $\varphi$  は  $C^\alpha$  原子の「前」の結合の二面角であるため、2 番目のアミノ酸の  $\varphi$  の  $b-c$  は「 $N_2-C_2^\alpha$ 」でなければならない。また、 $\psi$  は  $C^\alpha$  原子の「後」の結合の二面角であるから、2 番目のアミノ酸の  $\psi$  の  $b-c$  は「 $C_2^\alpha-C_2$ 」でなければならない。 $\omega$  はアミノ酸間の結合  $C-N$  の二面角であるため、2 番目のアミノ酸の  $\omega$  の  $b-c$  は「 $C_2-N_3$ 」か「 $C_1-N_2$ 」であろう。よって、これらの条件を満たすのは選択肢 1 であることがわかる。

問 55 正解【1】

PDB フォーマットは PDB の旧式の構造データファイルであり legacy PDB とも呼ばれる。PDBx/mmCIF は現在の PDB の正規フォーマットであり、PDBML はそこから XML 化により作成されたものである。一方 FAST は配列情報のフォーマットであり、立体構造情報の記述には適さない。よって選択肢 1 が正解である。

問 56 正解【1】

タンパク質の立体構造を PDBx/mmCIF 形式で記載したデータの読み方の問題である。loop\_の下に記述されたタグの順番に対応する様に、下方のデータの列を左から読めば良い。そのルールに基づき 15 番目のタグ\_atom\_site.B\_iso\_or\_equiv は、15 列目の温度因子に対応し、この列中での最大値は 18.86 である。同様に 6 番目のタグ\_atom\_site.label\_comp\_id に対応するのは、6 列目の VAL であり、同時に 8 列目の残基番号 11 が読み取れる。12 番目のタグ\_atom\_site.Cartn\_y に対応するのは 12 列目、原子座標の y 座標である。C<sup>β</sup>炭素は、4 列中の CB の行なので、これと対応する y 座標は 33.146 である。7 番目のタグ\_atom\_site\_label\_asym\_id と対応する 7 列目は、非対称単位中の分子の識別子（旧フォーマットのチェーンの識別子に相当する）A である。以上から、選択肢 1 が不適切であることがわかる。

問 57 正解【3】

図中の DNA は、右巻きであり、幅が広く深い溝（主溝）と、狭く浅い溝（副溝）が見られるので B 型 DNA である。図右側のタンパク質は β ストランドを多く含み、主溝と接していることがわかる。また図左側のタンパク質は、α ヘリックスの側面で主溝の部分と接している。接している領域には黒い棒球で示された Arg や Lys が集中していることが見て取れる。DNA の主鎖には負電荷のリン酸が含まれており、これに対応して正電荷を持つアミノ酸（Arg, Lys）が主に集中する。以上から、選択肢 3 が最も不適切であることがわかる。

問 58 正解【4】

並行 β シートが α ヘリックスで接続された β-α-β 構造は、TIM バレルやロスマンフォールドなどの α / β タンパク質で観察される超二次構造である。注意深く観察すると、(B)と(C)のそれぞれ N 末付近に、β-α-β 構造があることがわかる。図に 4 つの構造の PDB コードが記されているので、構造データをダウンロードし、分子グラフィックソフトを用いて目視で確認してもらいたい。

問 59 正解【4】

膜タンパク質は脂質二重膜を貫通して、両端が膜の両側に突き出ている。脂質二重膜の内部は疎水的な環境であるため、膜内で脂質分子と接するアミノ酸は、疎水性であることが多い。Ala, Leu, Ile, Val は疎水性なので、膜内に分布するはずである。一方、膜の両端に突き出た部分は、水分子に覆われているため、親水的なアミノ酸が多くみられる。ただし、細胞の外側と内側でアミノ酸組成は異なり、内側では、正の電荷をもったアミノ酸(Lys, Arg) がより多く見られる。この傾向は「ポジティブ・インサイド・ルール (positive inside rule)」と呼ばれ、膜タンパク質のトポロジー予測の基本原則である。これらを考えると、正解は選択肢 4 であることがわかる。ちなみに、図(A)は、芳香族アミノ酸 (Phe, Tyr, His, Trp) を彩色した図である。芳香族アミノ酸は、脂質二重膜と水の層の境界に現れる傾向がある。

問 60 正解【1】

ファンデルワールス相互作用と静電相互作用のポテンシャルエネルギーは、どちらも二原子間の距離によって定義される。ファンデルワールス相互作用の場合、距離の逆数の 12 乗と 6 乗の差で定義される。距離がゼロ付近の場合、12 乗の項が支配的になり、非常に大きな正の値になる。距離がある程度大きい場合は、6 乗の項が支配的になり、負の値から、ゼロに収束する。この間に極小値を持つ (詳細は平成 25 年度の間 55 を参照)。これを満たすのは左図の実線である。静電相互作用の場合、距離の反比例の曲線であり、極値は持たない。ただし、電荷の値の積  $q_1q_2$  の符号によって、値が負になるか、正になるかが変わる。積  $q_1q_2$  が正であれば左図の点線、負であれば右図の実線となる。よって、正解は選択肢 1 である。

問 61 正解【2】

アミノ酸配列上の特定の領域が  $\alpha$  ヘリックス、 $\beta$  ストランド、コイルの 3 状態のどれになるか予測することを 3 状態予測と言い、一般に  $\alpha$  ヘリックス状態予測の精度が最も良いと言われている。各状態に対し、他の状態とは関係なく別々に各アミノ酸の予測スコアが計算できるので、例えば  $\alpha$  ヘリックスと  $\beta$  ストランドの領域も一本のアミノ酸配列から独立に計算できる。PSIPRED 法は機械学習法のニューラルネットを用いた二次構造予測法であるが、PSI-BLAST で集めた関連ファミリー内での位置特異的スコア行列を学習セットにすることで予測精度を上げている。予測精度を評価するには、実際の立体構造上の二次構造と対応させる。DSSP は、立体構造上の主鎖間の水素結合から実際の二次構造領域を示せるのでこの目的で利用できる。以上から、選択子 1, 3, 4 は正しく、選択肢 2 が最も不適切である。

問 62 正解【1】

比較モデリング法の一つであるホモロジー・モデリング法において、鋳型として使える構造の配列類似性の下限は、だいたい 30%程度と言われている。配列類似度がもっと低くても立体構造が似ている場合もあるが、BLAST などの標準的な配列相同性検索では発見が困難となる。低い類似性の鋳型の発見には、プロフィール法、スレディング法などが有効であるとされる。これらも比較モデリング法に分類される。よって選択肢 1 が正解となる。

問 63 正解【1】

① オーソログとは、その系統関係が、生物種の系統関係を反映しているような遺伝子群のことであり、対応する（直系の）機能を担っていることが期待される。双方向ベストヒットの条件を課すことによって、生物種分岐の前に分岐していたタンパク質ペア（パラログ）をある程度排除できる。コドン使用頻度は生物種によって差があることが知られているが、オーソログの同定には役に立たない。

② 隠れマルコフモデル法を用いて、マルチプルアラインメントから、その族に属する遠縁のタンパク質を検索することができる（プロフィール検索）。Pfam データベースでは隠れマルコフモデル法のプログラム HMMer を用いている。細胞内局在は、遠縁のタンパク質間では必ずしも保存しないので、弱い相同性の認識には利用できない。

③ 膜貫通ヘリックスの予測は、膜タンパク質すべてに共通する傾向を予測原理とするため、相同蛋白質の情報がなくともある程度予測できる。分子ドッキング法は、二つの分子の単体の立体構造のデータがなければ適用できないので、ここでは不適當である。これらを考え合わせると、正解は選択肢 1 である。

問 64 正解【1】

ツーハイブリッド法としては、出芽酵母を用い、GAL4 という転写因子を用いた方法が最初に提案された。この方法は、特に酵母ツーハイブリッド法 (Yeast two-hybrid; Y2H) と呼ばれる。免疫蛍光法とは、抗体を蛍光色素で標識することで、その抗体が結合する分子（抗原）が分布する領域を発色させ、蛍光顕微鏡で観察する方法である。これはタンパク質間相互作用の計測法ではなく、細胞標本のイメージング手法の一つである。よって選択肢 1 が正解となる。

問 65 正解【3】

ヒトのミトコンドリア・大腸菌・高度好熱菌のなかで、ユニバーサルコドンを利用してないのは、ヒトのミトコンドリア。また、高度好熱菌は大腸菌と比較すると、GC が豊富なコドンを使う傾向にある。

問 66 正解【1】

遺伝子の多重化（重複）に、その遺伝子の進化速度が上がるということが直接関与するとは考えにくい。

問 67 正解【2】

Ser と Arg はコドンの 1 文字目や 2 文字目の変化においても同義である。

問 68 正解【1】

ハプロタイプを決定するには、各遺伝子型とその組み合わせがわからなければならない。

問 69 正解【2】

ヒトとチンパンジーの間で完全に保存されている塩基部位は、3 箇所である。

問 70 正解【2】

世代時間が分子時計の速さに影響するという一般的事実は、知られていない。

問 71 正解【2】

分子系統樹の作成方法としては、最尤法、近隣接合法、および最大節約法が有名である。ウォーターシェッド法は、データの分類などの際に用いる領域分割方法のひとつ。

問 72 正解【1】

系統樹の両枝に、異なる生物種のオーソログが含まれていることより、 $\alpha$ -グロビンと  $\beta$ -グロビンの分岐は、種分岐以前であることがわかる。

問 73 正解【3】

タンパク質の立体構造は物理化学的性質に基づく多体問題であるため、核酸配列との単純な対応付けはできない。

問 74 正解【2】

一般に、量的形質は質的形質よりも環境の影響も受けやすい。逆に環境に影響を受ける形質は量的形質であることが多いとも言える。これは遺伝型は離散的な特性であるが、環境は連続的な特性であることから理解できる。

問 75 正解【4】

タンパク質間相互作用ネットワークは、ノードの次数がべき乗分布となるスケールフリーネットワークであることが知られている。

問 76 正解【1】

解析対象生物種のゲノム配列から、その生物種が持ちうるタンパク質配列を列挙することが可能である。トリプシンはリジン。アルギニンのカルボキシル基側のペプチド結合を加水分解する特性があることから。各タンパク質をトリプシン分解した際のペプチド質量が推定できる。予め推定ペプチド質量のデータベースを作成しておき、実際に得られた質量と比較することでタンパク質同定が可能になる。

問 77 正解【3】

完全グラフは全てのノードが互いに連結しているグラフのことを指す。任意のタンパク質間に相互作用が存在するのではないため、タンパク質間相互作用ネットワークは完全グラフではない。

問 78 正解【1】

質量分析装置は一般に小分子の測定に適している。ゲノム配列は非常に長いいため、アミノ酸配列の決定と同様の手法では配列決定することができない。

問 79 正解【4】

NMR は質量分析器より感度が低いものの、原理的には一度に全メタボライトを解析できる網羅性を持つ。

問 80 正解【3】

他の遺伝子の操作により遺伝子 D の発現量が変化する実験結果は得られていないため、遺伝子 B が遺伝子 D を制御していると推定する根拠はない。

試験問題に記載されている会社名または製品名は、それぞれ各社の商標または登録商標です。  
なお、試験問題では、®および™を明記していません。

Copyright © 2014 Japanese Society for Bioinformatics. All Rights Reserved.