

2019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【4】

模式図の特徴から (ア) 核、(イ) 小胞体、(ウ) ゴルジ体、(エ) ミトコンドリア、(オ) リソソームである。

問2 正解【4】

核小体は膜を持たない核内の構造体で、rRNA の合成とリボソームの産生のもととされている。

問3 正解【3】

細胞分裂は、DNA 合成の準備期間である G1 期から、DNA の複製が行なわれる S 期へと進み、G2 期は分裂 (M 期) の準備期間である。よって G2 期に細胞に含まれる DNA の量は G1 期の 2 倍である。

問4 正解【3】

一般に、mRNA の 3' 末端にポリ A 鎖が付加されるのは真核生物である。

問5 正解【2】

一般に、真核生物の細胞質のリボソームには 5S rRNA、5.8S rRNA、18S rRNA および 28S rRNA が含まれており、原核生物 (真正細菌と古細菌) のリボソームには 5S rRNA、16S rRNA および 23S rRNA が含まれている。

問6 正解【2】

ジスルフィド結合ではなく、ホスホジエステル結合が正しい。ヌクレオチドの五炭糖の 3' 炭素に共有結合したヒドロキシ基 (-OH) が (5' 炭素に共有結合した) リン酸基とホスホジエステル結合することにより重合 (高分子化) したものを核酸とよぶ。なお、同じタンパク質内あるいは異なるタンパク質間において、2 つのシステイン残基の側鎖のチオール基 (-SH) がジスルフィド結合する場合がある。

問7 正解【4】

グリシン以外の 19 種類のアミノ酸は側鎖に炭素原子を含むが、グリシンの側鎖には 1 個の水素原子しか含まれない。

問8 正解【1】

一般に、イオンを輸送する膜タンパク質であるイオンチャネルは 2 回以上の膜貫通部位を持ち、配列だけでなく膜貫通の回数やトポロジーでファミリーとして分類されることがある。

問9 正解【2】

タンパク質への糖鎖修飾には N-グリコシル化と O-グリコシル化の二種類がある。ユビキチンは 76 残基程度の小型のタンパク質で、ポリユビキチン修飾されたタンパク質はプロテオームにより認識され分解を受ける。

問10 正解【2】

解糖系はグルコースの分解に酸素を消費しない嫌気的な経路である。

問11 正解【2】

(A)は近傍の標的細胞に作用するパラクリン型、(B)は軸索を通じて電気シグナルが伝えられる神経型、(C)はシグナル伝達物質であるホルモンが血管を通じて標的細胞に到達する内分泌型、(D)は細胞表面に存在する膜タンパク質が標的細胞の受容体に直接作用する接触型、それぞれの模式図。

問12 正解【1】

R と r それぞれの対立遺伝子がホモ接合の個体を交配した F1 では、ヘテロ接合の個体のみが得られるが、F2 では配偶子の組み合わせから、遺伝子型では RR (丸)、Rr (丸)、rr (しわ) の個体がそれぞれ 1:2:1 の頻度で得られ、表現型では顕性:潜性 3:1 の頻度となる。

問13 正解【3】

一般に、哺乳類の種ではミトコンドリアとそのゲノムは母親から子供に伝わる。父親から子供にミトコンドリアが伝わったという論文報告もあるが、検証が必要な段階である。

問14 正解【3】

G はアミノ酸だとグリシンの 1 文字表記である。グリセリンは 3 価のアルコールであるグリセロール (1,2,3-プロパントリオール) の別名 (旧称)。なお、国際塩基配列データベースの核酸コードでは、塩基の略号は小文字 (a, c, g, t) を使用することになっている。ただし、RNA 配列の場合、ウラシルは u ではなく t (チミンと同じ略号) を用いる。

問15 正解【4】

LINE は長鎖散在反復配列(long interspersed nuclear element)の略称。ヒト遺伝子のプロモータ領域に存在する GC 含量の高い領域にある反復配列は CpG アイランドとよばれる。

問16 正解【2】

PCR は変性、アニーリング、伸長の 3 ステップにおいて、反応温度を変化させて増幅を行う。

問 17 正解【4】

ゲノム DNA から主にエキソン領域の配列決定を行う NGS 手法は一般にエキソーム解析とよばれている。メタボローム解析は質量分析装置による代謝物の網羅的解析手法。

問 18 正解【3】

CRISPR/Cas9 によるゲノム編集手法は、現在、大腸菌から植物、ヒトまで幅広い生物種において有効な実験手法となっている。

問 19 正解【1】

タンパク質の立体構造決定法は、試料の状態や分解能、分子量限界により、それぞれ長所と短所がある。分子量限界は NMR 法が最も小さく、次に X 線結晶解析、電子顕微鏡の順である。

問 20 正解【3】

「ヒトゲノム・遺伝子解析研究に関する倫理指針」の基本方針には、『指針に基づく研究計画の作成及び遵守並びに独立の立場に立った倫理審査委員会による事前の審査及び承認による研究の適正の確保』が掲げられている。倫理審査委員会は『独立の立場に立って、学際的かつ多面的な視点から、様々な立場からの委員によって、公正かつ中立的な審査を行えるよう、適切に構成し運営されなければならない。』とされており、研究者以外の倫理・法律の専門家や一般の立場の委員で構成される。申請書の作成や承認には、多くの労力と期間が必要とされている。

問 21 正解【3】

排他的論理和は二つの値が等しい時に 0、そうでないときに 1 の値をとる。この選択肢においては、 $\text{int}(t(x) \text{ XOR } t(y))$ が $(x+y)$ を 2 でわった余りとなることに気づくと、3 が正解であることがわかる。効率よく回答するには、適当に数値をいくつか入れてみて解くこともできるかもしれない。

問 22 正解【4】

GPGPU は、もともと画像処理を高速化するための演算ユニットを一般計算目的にも用いることが可能なように拡張した演算ユニットのことである。

問 23 正解【2】

通常、公開鍵暗号は、公開した公開鍵で暗号化し、秘匿している秘密鍵で復号する。RSA 暗号は素因数分解問題、ElGamal 暗号は離散対数問題とよばれる問題と関係がある。

問 24 正解【2】

キーの衝突は、ハッシュ関数の値が一致する ($h(x_1) = h(x_2) = r$) ときに起こる。このとき、 x_1 と x_2 をそれぞれ n で割ったときの商を q_1, q_2 とすると、

$$x_1 = q_1 \times n + r$$

$$x_2 = q_2 \times n + r$$

となる。2つの式の差をとると、 $x_1 - x_2 = (q_1 - q_2) \times n$ となるので、 $x_1 - x_2$ が n の倍数のときにキー x_1 とキー x_2 が衝突する。

問 25 正解【4】

あるグラフの隣接行列とは、そのグラフ上で辺で結ばれている頂点の組み合わせに対応する行列要素が 1、それ以外が 0 となっている行列のことを言う。

問 26 正解【1】

入力がソートされているため while ループの内部は実行されず、 $\Theta(n)$ 時間で終了する。挿入ソートの最悪計算量は $\Theta(n^2)$ だが、この例のように入力がすでに殆どソートされている場合など、高速に動作する場合もある。

問 27 正解【2】

入力サイズが n の時の実行時間を $T(n)$ とすると、 $T(n)$ は適当な定数 $c > 0$ に対して

$$T(n) = \begin{cases} O(1) & \text{if } n \leq c \\ T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil) + \Theta(n) & \text{if } n > c \end{cases}$$

を満たす。

問 28 正解【1】

上位 5%未満の値を返す確率は 0.95^m であり、これが 1%以下となるのは $0.95^m < 0.01 \Leftrightarrow m > 20 \ln(100) \approx 92.1$ の時である。求める m の値は n にはよらない。

問 29 正解【4】

この図はソーティングネットワークとよばれる図の一例となっており、E, F, G, H は A, B, C, D がソートされたものとなる。ソーティングネットワークは並列計算機上でのソートアルゴリズムの一種であり、また、入力によって計算時間が変わることのないアルゴリズムとしても知られている。

問 30 正解【3】

最短経路の数は、5つの部分経路の候補のうち、どの2つを垂直部分に選ぶかに相当し、 $\binom{5}{2} = 10$ 通りとなる。

問 31 正解【1】

どの最短経路についても、右方向に向かう水平部分と、下方向に向かう垂直部分から構成されるため、選択肢1が正しい。

問 32 正解【1】

SQL はリレーショナルデータベースの検索のための標準言語である。

問 33 正解【1】

前半の SELECT 文によって、「札幌商品」表から「東京商品」表に含まれる商品を除外した表が得られる。これを関係代数式で表すと、(札幌商品 - 東京商品) となる。同様に後半の SELECT 文によって、(福岡商品 - 東京商品) が得られる。これらを UNION で和演算することで、(札幌商品 - 東京商品) \cup (福岡商品 - 東京商品)、すなわち選択肢1の (札幌商品 \cup 福岡商品) - 東京商品が得られる。

問 34 正解【2】

正規分布は左右対称な分布である。

問 35 正解【1】

ある分布において、ある区間に落ちる確率を計算するにはその区間の積分値を計算すればよいが、ある分布の累積密度関数はその分布を積分したものである。

問 36 正解【2】

この標本の場合、 $r_{xy} = -0.852$, $r_{xz} = -0.943$ であった。図より、 $r_{xy} < 0$ および $r_{xz} < 0$ であり、これを満たす選択肢は2しかない。

問 37 正解【3】

最急勾配法は最も値が小さくなる方向に向かって局所最適解をめざす方法である。なお、得られた局所最適解はあくまでその領域における最適解であり、全体の最適解（最小解）であるとは限らないことに注意が必要である。

問 38 正解【4】

入力ベクトル $\mathbf{x}_a, \mathbf{x}_b \in R^d$ がいずれも陰性と予測されるということは、 $\mathbf{w}^\top \mathbf{x}_a + b < 0$ 、かつ $\mathbf{w}^\top \mathbf{x}_b + b < 0$ がいえる。この事実を使う、 \mathbf{x}_c のスコアは

$$\mathbf{w}^\top \mathbf{x}_c + b = \mathbf{w}^\top \left(\frac{\mathbf{x}_a + \mathbf{x}_b}{2} \right) + b = \frac{1}{2}(\mathbf{w}^\top \mathbf{x}_a + b) + \frac{1}{2}(\mathbf{w}^\top \mathbf{x}_b + b) < 0$$

のように負の値になることがわかる。よって、 \mathbf{x}_c は陰性と予測される。

問 39 正解【1】

隠れマルコフモデルは学習方法の一つではあるが、クラスタリングの手法とは一般的には言えない。UPGMA は系統樹の作成などにも用いられる。

問 40 正解【4】

n -fold 法とは、データの $1/n$ をテストに用い、残りを訓練に用いるクロスバリデーションのことである。

問 41 正解【3】

メタゲノムデータの Binning においては、同一ゲノム内では全長にわたって概ね均一で、かつゲノム毎には異なる値をとるような特徴量を用いるのが効果的である。選択肢 1 の k -mer 頻度はそうした性質を持つことが知られており、また選択肢 4 の被覆度(coverage)も、環境中でそのゲノムを持つ生物の存在量を反映するため、そうした性質を持つ。これに対し選択肢 3 の遺伝子機能カテゴリの割合は、ゲノム全体で均一ではなく、ゲノム毎の違いもさほど明確でないため、有効ではない。よってこれが正解。選択肢 2 の単一コピー遺伝子は、すべてのゲノムに共通という点で上述の特徴にはあてはまらないが、「ゲノム中ただ 1 コピー存在する」という条件を制約として扱うことで、Binning の結果を評価するのに用いることができる。

問 42 正解【4】

接尾辞配列は、与えられた文字列の接尾辞を辞書順にソートしたときの、各接尾辞の元の文字列における出現位置の配列のこと。左右を見比べて、左端の出現位置を、右側の対応する接尾辞の順に沿って並べ替えればよい。

問 43 正解【4】

GenBank 形式において、ORIGIN 行は配列データ開始の直前に置かれる目印として知られる。ここには何も記載されていないので、選択肢 4 が正解。もともと ORIGIN 行は、配列データの先頭がゲノム上のどこに位置するかに関する情報を記載するフィールドであり、かつてはここに選択肢 4 のような情報が記載される可能性はあったが、現在では基本的に

ORIGIN 行は空白で、もっぱら配列データ開始の目印として用いられている。なお、FEATURE 行に rep_origin というキーワードが定義されており、複製開始点の情報が記載される場合はそこに記載される。

問 44 正解【3】

配列内に繰り返し配列があると、ドットプロットには、繰り返し配列間で一致する領域すべてがプロットされ、選択肢 4 のように複数の平行線が生じる。よって、選択肢 3 は選択肢 4 を左右反転させたプロットになるはずである。なお、問題の配列 tgca は、逆相補配列が自身と一致する回文配列 palindrome なので、相補配列として比較すると逆向きの対角線がプロットに表れる。

問 45 正解【2】

アラインメント中に、一致が 4 つ、不一致が 1 つ、ギャップが 1 つ含まれるので $3 \times 4 + 1 \times 1 + 2 \times 1 = 9$ で答えは 9 となる。なお、「ギャップペナルティ」は正の数値で表すことが多いが、その分だけマイナスするという意味になるので、間違えないように。

問 46 正解【2】

図中の四角は、代謝の過程の中間生産物ではなく酵素を表し、中の数字は対応する酵素番号である。よって選択肢 2 が誤り。なお、選択肢 3 の例は、図の右下に spontaneous と記載された矢印として表されている。

問 47 正解【3】

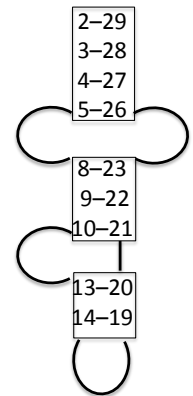
k -mer をノードとするド・ブラングラフにおいては、矢印の前後で隣接するノードは、その順に並べたときに互いに $k-1$ 文字の重複を持つ文字列である。従って(a)は CC で始まり、(c)は GC で終わることが直ちに分かる。このように考えて穴埋めしていくと、選択肢 3 が正解と分かる。

問 48 正解【1】

RNA-Seq 解析においては、シーケンサから得られた RNA 由来の配列データ（リード）を、その生物のゲノム配列が知られている場合はそれにマッピングし、そうでない場合は新規にアセンブルすることで、遺伝子配列を得ることができる。いずれの場合も、リードを遺伝子配列にマッピングして数えることにより、発現量解析が行われる。リード数の差は発現量を反映するため、繰り返し配列の検出には向かない。よって選択肢 1 が誤り。なお、CDS はタンパク質をコードする配列(coding sequence)のことで、アセンブリで得られた mRNA 配列から推定できる。

問 49 正解【4】

対応する塩基対において、左側の数値は単調に増加し、右側の数値は単調に減少しているため、これらは直鎖状の二次構造を形成している(右図参照)。塩基 2 から 5 まではステム、塩基 11 から 12 までは左側のみでループを形成しているためバルジループ、塩基 15 から 18 まではステムの端で折り返すヘアピンループに相当する。この直鎖状二次構造において、塩基対が交差することなく表されるためシュードノットは存在しない。よって選択肢 4 が誤り。



問 50 正解【1】

配列比較に用いられる代表的なアミノ酸置換行列である BLOSUM や PAM は、実際の相同配列間のアラインメント中に観測されるアミノ酸置換の頻度に基づいて作成されている。物理化学的性質が似たアミノ酸は置換されやすい性質を持つため、それがスコアにも反映されているが、直接的に物理化学的指標を用いて定義されたわけではない。よって選択肢 1 が正解。なお、PAM の数値は、その行列を定義する際に想定した配列間の進化距離を表しており、アミノ酸配列 100 残基あたりに生じた置換の回数と定義される。従って、数値が大きい方が進化的に遠い配列比較に向いている。BLOSUM は数値が小さい方が進化的に遠い配列比較に向いているため、逆になる点に注意。

問 51 正解【2】

ゲノムが十分に長い(末端の影響が無視できる)とき

$$P(G) = P(GA) + P(GC) + P(GG) + P(GT) = P(AG) + P(CG) + P(GG) + P(TG)$$

が成り立つ。どちらを計算しても $P(G) = 0.2$ となる。

後半は、条件付き確率の式に当てはめると、

$$P(T|G) = P(GT) / P(G) = 0.05 / 0.2 = 0.25 \text{ となる。}$$

条件付き確率 $P(T|G)$ (=1 塩基目が G である条件のもとで 2 塩基目が T である確率) を、 $P(GT)$ (=1 塩基目が G かつ 2 塩基目が T である確率) と混同しないように。

問 52 正解【3】

問題文中の式に当てはめると

$$P(GTC) = P(G)P(T|G)P(C|T)$$

$P(G)$ と $P(T|G)$ は問 51 で計算しているため、あとは $P(C|T)$ を計算すればよい。これも問 51 と同様にして計算できる。

$$P(C|T) = P(TC) / P(T) = 0.06 / (0.07 + 0.06 + 0.07 + 0.1) = 0.2$$

$$P(GTC) = 0.2 * 0.25 * 0.2 = 0.01$$

問 53 正解【3】

続く問 54 と共に、タンパク質の構造を格子上で表現したモデルを用いて考察する問題である。タンパク質フォールディングの理論研究において、三次元格子モデルは重要な役割を果たしてきた。今回は問題を単純にするため二次元格子を用いているが、基本的な考え方は共通している点が多い。

問題文に示されているアミノ酸配列を格子上に当てはめると、溶媒に露出している部分と埋没するアミノ酸残基が明らかになる。このモデルでは埋没している（上下左右が他のアミノ酸残基に占められている）部分は中心の二残基のみであり、ここに疎水性残基（I と V）が位置するようなフォールドが答えとなる。

問 54 正解【1】

アミノ酸共進化とは何かを詳しく述べることは簡単ではないが、極端な例をひとつ挙げると下記のように説明できる。例えば 2 つのアミノ酸残基、K と D が塩橋を形成して機能上重要な役割を果たしているとする、塩橋を崩すような変異は進化上起こりにくいだろう。一方、これら 2 つのアミノ酸の位置が入れ替わるような二重変異であれば塩橋を維持する可能性があるため、比較的起こりやすいものと期待できる。つまり、構造上の特徴によって 2 つのアミノ酸座位での変異の傾向が相関することになる。マルチプルアラインメントからこのような相関を見つけ出し、構造上の特徴（主に残基間接触）を予測することが可能である。

問題文中のマルチプルアラインメントでは、2 行目から 3 行目の間で 2 番目と 11 番目のアミノ酸残基（K と D）が入れ替わっている。また 4 行目から 3 行目の間で 3 番目と 10 番目のアミノ酸残基（L と V）が入れ替わっている。従って、これら 2 対のアミノ酸（2 と 11、および 3 と 10）が近接しているフォールドが答えとなる。

これはあくまで考え方を単純なモデルに落とし込んだものであり、実際にはもっと複雑である。詳細については各方法論の原著論文を当たるのが良いだろう。

問 55 正解【3】

蛋白質立体構造データベース（PDB）のデータフォーマットに関する問題である。問題文中にもあるとおり、データ登録については従来のいわゆる PDB フォーマットはすでに廃止され、新しい規格である mmCIF フォーマットによる登録が義務づけられている。mmCIF フォーマットの特徴を理解するにあたって、何故 mmCIF フォーマットが必要なのか、PDB フォーマットでは何ができないのかを把握しておくことは有益だろう。また、PDB フォーマットは非常に手軽に扱えることから、PDB のデータ登録以外の場面では依然としてよく使われており、その特徴を理解しておくことも重要である。

選択肢 3 は、人間がファイルを開いて見た場合の視認性の問題であり、プログラムによる判別では問題にならない。

問 56 正解【4】

図示されているのは B 型 DNA である。B 型 DNA が一般的だが、他にも幾つかの例外的な型の構造が報告されている。B 型 DNA には以下の特徴がある。二重らせんは 2 本のポリヌクレオチド鎖で形成され、それぞれの鎖は逆方向を向いている。形成されるらせん構造は右巻きである。相補的塩基対は水素結合により結合している。二重らせんは約 10 塩基対で一回転し、主溝 (major groove) と副溝 (minor groove) よばれる溝を形成する。図示された DNA が右巻きであることは、手前側に描かれた鎖が右上と左下を結んでいることでわかる。ウェブ上から DNA のフリー素材を取得する際など気をつけられたい。

問 57 正解【4】

タンパク質立体構造を中心とした統合データベースの一つである PDBsum からの情報の読み取りを扱った問題である。配列、立体構造、様々な機能の情報を取得できるため、解析対象タンパク質について調査するときには有用である。問題紙面の都合上、限られた情報のみ示したが、この他にも様々な情報が提供されているため、一度利用してみることをおすすめする。選択肢 1 から 3 の内容は図中に記載されているため、落ち着いて読み取ることができていれば正答できただろう。CATH データベース由来の立体構造分類情報は Mainly Beta Beta Barrel と図中に記載されている。したがって、選択肢 4 が不適切である。

問 58 正解【4】

異なる祖先型タンパク質どうしが、類似機能をもつように進化的に収斂する場合がある。このような関係にあるタンパク質間では、機能の類似性のみを示し、配列間および立体構造間のいずれにおいても類似性を示さない。このように、異なる配列、立体構造を有するタンパク質が類似機能へと進化的に収斂する例が多く報告されている。選択肢 1 から 3 はいずれも収斂進化によるタンパク質機能の類似性に関連した記述である。

問 59 正解【2】

ハブタンパク質は複数の相手と相互作用することが知られているが、相互作用部位および相互作用のタイミングは多様で、相互作用部位が同じでも相互作用するタンパク質が同じであるとは限らず、同時に複数の相手と相互作用することがある。立体構造的な特徴についても様々な情報が得られているが、天然変性領域を持つことで多種の相手と相互作用することが知られている。

問 60 正解【3】

タンパク質と基質が結合する際のモデルとして、構造変化を考慮しない「鍵と鍵穴(lock-and-key) モデル」と、基質がタンパク質に接近することでタンパク質の構造変化が引き

起こされる「induced-fit モデル」、およびタンパク質の構造変化揺らぎが基質の接近にすることによって結合型の構造になる「pre-existing モデル」の三つがある。

問 61 正解【1】

塩基置換や欠失が有意に多いことから、ゲノムの他の箇所から読まれたリードが、図に示したリファレンスゲノムの箇所に誤ってマッピングされたと想定される。これは、リードが読み取った配列がリファレンスゲノムに存在しない場合に起こり得る。選択肢 2(重複箇所がリファレンスゲノムに存在しない)、3(リファレンスゲノムに未決定箇所がある)、4(他生物の配列が混入している)が、この場合に適合する。従って選択肢 1 が正解となる。

問 62 正解【3】

LINE や LTR トランスポゾンなどの長い反復配列は、レトロウイルスゲノムに類似した構造を持ち逆転写酵素をコードしている。一方、SINE に属する Alu などの短い反復配列は ORF を持たない。よって選択肢 3 が正解となる。

問 63 正解【2】

水平伝播した遺伝子では、塩基配列などで推定した分子系統樹が種系統樹と大きく異なることが多い。また水平伝播した遺伝子では、GC 含量、コドン使用頻度、 k -mer(塩基数個の配列)頻度などが伝播元の生物の特徴をまだ維持していて、伝播先の生物とは大きく異なる場合が多い。一方、タンパク質の構造モチーフは、水平伝播した遺伝子の機能を反映しており、種系統やコドン使用頻度などとの相関は低い。よって選択肢 2 が正解となる。

問 64 正解【1】

主として連鎖不平衡は、同一染色体上に近接して存在し遺伝的連鎖する複数の遺伝子座の特定の組み合わせ(ハプロタイプ)にランダムではない相関が見られる現象であるが、例外的に異なる染色体に存在する遺伝子座の間でも有意な相関を示すことが知られている。よって選択肢 1 が正解となる。

問 65 正解【1】

図の系統樹から、ノード 1 が α -グロビンと β -グロビンの遺伝子重複(の最終祖先)に相当し、ノード 2 およびノード 4 以降が種分岐に相当することがわかる。従って、ノード 2 またはノード 4 以降に連なる遺伝子(タンパク質)はそれぞれオーソログの関係にあり、その他の遺伝子(タンパク質)の組み合わせはパラログの関係にある。よって選択肢 1 が正解となる。

問 66 正解【3】

図の系統樹に 1 つの OUT を加えることは、分岐させる枝を現状の枝から 1 つ選択することに相当する。従って、{現状の系統樹の可能な樹型の数}×{現状の枝の数}、すなわち $15 \times 7 = 105$ が求める可能な樹型の数となる。よって選択肢 3 が正解となる。

問 67 正解【4】

自然選択に対して有利でも不利でもない遺伝子変異とは、機能的な制約が無いか非常に低いことを意味する。発現しない偽遺伝子、アミノ酸配列を変化させない同義置換、アミノ酸配列を変化させない可能性が高いコドンの 3 番目の変異がこれに相当する。一方、加速進化は自然選択により有利な変異が蓄積することを意味する。よって選択肢 4 が正解となる。

問 68 正解【4】

$-\log_{10}p$ 値は選択肢 2 の危険率(p 値)から求められる値である。一方、オッズ比は、以下の表で $a \sim d$ を観測された人数として $(a/b)/(c/d)$ で求められる値であり、 $-\log_{10}p$ 値とは異なる。また、 $-\log_{10}p$ 値は高いほうが疾患リスクが高いことを意味する。よって選択肢 4 が正解となる。

	疾患群	対照群
SNP あり	a	b
SNP なし	c	d

問 69 正解【2】

1 階の常微分方程式を差分方程式に置き換え、初期値から逐次更新で数値積分を行うオイラー法を用いる問題。微分方程式を差分方程式 $(X_{t+\Delta t} - X_t) / \Delta t = k X_t$ とおきかえ整理すると、 $X_{t+\Delta t} = (1 + k\Delta t) X_t$ となり、従って $X_2 = (1 + k\Delta t)^2 X_0 = (1 + 0.1)^2 \cdot 1 = 1.21$ となる

問 70 正解【1】

A は自身への活性化があるため、自己触媒的に増えるが、B に関しては A との関係性のみ定義されており、逐次的な時間変化の定性的な振る舞いとしては B の自己触媒的記述は不適切。A, B それぞれの間のみの関係性についての、時間変化の A, B 値依存性 (時間発展方程式におけるヤコビアン) の正負) と関係する問いである。

問 71 正解【1】

固有値は 1, 2 であり、すべての固有値の実部が負ではないため、システムの平衡点まわりの挙動は不安定である。

問 72 正解【2】

通常、DNA のメチル化はシトシン (C) で起こり、遺伝子の発現は抑制される。

問 73 正解【4】

D が終止コドンであることは間違っていないが、終止コドンはゲノム上の配列としては TAA, TAG, TGA の 3 種類であることが知られている。(2021.1.13 修正)

問 74 正解【4】

質量分析装置は、様々な物質の質量を測定する装置であり、様々な種類のクロマトグラフと組み合わせて物質の分析に用いられる。多段の質量分析などにより、タンパク質のアミノ酸配列までは同定できることもあるが、立体構造の決定には用いられていない。

問 75 正解【2】

同じ試料で細菌の存在量の比較では規格化は必要ないが、異なる試料間では総リード数の違いと細菌数の違いを考慮して、総リード数で割って細菌数を掛ける規格化が必要となる。

問 76 正解【3】

メタゲノム解析やメタ 16S 解析で取得した細菌叢データの分析において、とある環境の種多様性を表す α 多様性 (α -diversity) と、環境間の種多様性の違いを表す β 多様性 (β -diversity) が議論されることが多い。 α 多様性としては、Shannon index や Simpson index がよく用いられ、 β 多様性としては、UniFrac distance が広く用いられる。なお、UniFrac distance は含まれる細菌の量も考慮して計算する場合 (weighted) と、考慮しないで存在の有無だけで計算する場合 (unweighted) がある。

問 77 正解【4】

タンパク質間相互作用の結果できる大きなタンパク質複合体の場合には、全ての要素タンパク質の間に直接の相互作用が無い場合もあり完全グラフにはならない。

問 78 正解【4】

様々な理由で mRNA とそれが翻訳されて合成されるタンパク質の量比は遺伝子毎に異なる。主な理由は 1-3 のような場合である。mRNA の転写速度は mRNA の量には影響し、翻訳速度はその mRNA に対応するタンパク質の量に影響する。

問 79 正解【3】

モノクローナル抗体は特定のエピトープを認識し分子種が均一であるが、ポリクローナル抗体はいろいろな抗体分子種の混合物である。そのため、エピトープが化学修飾やスプライ

シング、プロテアーゼ消化などによって失われると、モノクローナル抗体は標的タンパク質を認識出来なくなる。

問 80 正解【3】

ベイジアンネットワークでは、サイクルを持つネットワークを扱わないので、(3)以外の非サイクル (acyclic) なネットワークが推定結果となる。

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。