

平成27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【2】

原核生物のゲノム DNA は多くの場合環状である。

問 2 正解【4】

ウイルスは宿主細胞の機能を乗っ取って自身の複製を作るための部品を大量に合成する。合成された部品が組み立てられて新しいウイルスが大量に作られ細胞外へと放出される。

問 3 正解【2】

解糖系は細胞質、クエン酸回路（TCA 回路）と電子伝達系はミトコンドリアに存在する。

問 4 正解【2】

ミトコンドリアは環状のゲノム DNA を持つ。

問 5 正解【4】

問題文から②が M 期であることが判るので、③が G<sub>1</sub> 期、④が S 期、①が G<sub>2</sub> 期である。この時点で選択肢 2 と 3 は正答から除外される。S 期にゲノムは複製されるため、その直後の G<sub>2</sub> 期では DNA 量が 4n となるはずであるから、選択肢 1 も間違いである。

問 6 正解【2】

減数分裂では父方由来と母方由来の染色体を持つ 2 倍体の細胞（2n）から 1 倍体（1n）の細胞が生成する。減数分裂に先だってゲノムが複製されて倍加し（4n）、ついで減数第一分裂と減数第二分裂の 2 回の分裂を経て 1 倍体（n）の配偶子が形成される。

問 7 正解【2】

AaBB の親から生じる配偶子は AB と aB の遺伝型を持つ。AABb からは AB と Ab が生じる。これらの組み合わせを考えると良い。

問 8 正解【2】

1 つの B 細胞が産生する抗体の種類は 1 種類である。多様な抗原へ対応するためには多くの B 細胞（とそれらが産生する抗体）が必要になる。

問 9 正解【1】

特になし

問 10 正解【1】

特になし

問 11 正解【4】

A と T の間には 2 つの、G と C の間には 3 つの水素結合が形成されている。水素結合の数が多いため G と C の対合は A と T の対合に比べて安定である。

問 12 正解【3】

複製フォークの進行方向と同一方向に合成される鎖をリーディング鎖、逆方向に合成される鎖をラギング鎖と呼ぶ。リーディング鎖では DNA が連続して合成され、ラギング鎖では断続的な合成とそれらの連結が行われる。

問 13 正解【2】

RNA はチミンの代わりにウラシルが利用されている。

問 14 正解【3】

アミノ酸には L 体と D 体が存在し、生体内では L 体の存在比が圧倒的に高い。化学合成されたアミノ酸には L 体と D 体がほぼ等量含まれることがあり、こういった混合物をラセミ体という。

問 15 正解【4】

終止コドンに結合するのは翻訳終結因子と呼ばれるタンパク質である。

問 16 正解【3】

プロテアソームはタンパク質の分解を行う複合体である。

問 17 正解【3】

真核生物の核内の DNA はヒストン 8 量体が巻き付いたヌクレオソーム構造を形成しているが、ミトコンドリアや葉緑体のゲノム DNA はヌクレオソーム構造を形成しない。

問 18 正解【3】

PCR 反応は一般的に反応液の「温度」を変化させることで変性、アニーリング、伸長のそれぞれのステップを行う。

問 19 正解【3】

細菌叢から DNA を抽出し、配列決定と解析を行う手法はメタゲノム解析である。

問 20 正解【1】

特になし

問 21 正解【3】

仮想記憶が技術の名称。ページング方式の仮想記憶では必要に応じて二次記憶装置上のスワップファイルもしくはスワップパーティションにメモリの内容を書き出し、主記憶を他のプロセスが使えるようにする。

問 22 全員正解としました。

本来はパターン完全一致を意図した問題でしたが、出題意図が十分に伝わらないという指摘があり、検討を重ねた結果この問題は全員正解としました。すなわち、Perl の場合 `/^J[a-z]*[0-9]*-[a-z]*$/` でしたら、2 が正解になりますが、`^` および `$` がついていないので、すべての選択肢がマッチしてしまいます。

問 23 正解【1】

まず、 $(A, B, C) = (0, 0, 1)$  のとき、A と B の OR をとると 0 となり、その結果と C の AND をとると 0 となる。次に、 $(A, B, C) = (1, 1, 0)$  のとき、A と B の OR をとると 1 となり、その結果と C の AND をとると 0 となる。

問 24 正解【3】

クラウドコンピューティングでは既に必要なソフトウェアのインストールが行われたシステムを借りることで、ソフトウェアのインストール作業を改めて行う必要性から解放されることが利点の一つなので、選択肢 3 は誤りである。

問 25 正解【3】

図 1 は perl によって書かれた 2 つの数の足し算を行うサブルーチンである。この言語はインタプリタ言語の一種であり、コードの 1 行 1 行が計算中に評価される。一方図 2 は C 言語によって書かれた整数の和を求めるサブルーチンである。これはコンパイラ言語の一種であり、事前にコンパイルし機械語に変換された後、計算が行われる。

問 26 正解【1】

TCP/IP はネットワーク上で異なる OS 間などでも通信することを可能にしたネットワークのプロトコルである。

問 27 正解【2】

数値を順番にスタックに push し演算子が来たらスタックから値 2 つを pop して、演算結果を push すれば実装できる。選択肢 1, 3, 4 のデータ構造を用いて同じ演算を実装することは可能であるが、最適とはいえず一般的ではない。

問 28 正解【2】

二分探索木では、各節点について、節点の左の子孫はすべてその節点よりも小さく、右の子孫はすべてその節点よりも大きくなっている（ただし、大小関係はその逆でもよい）。二分探索木では、任意のラベルを持つ節点をその深さに比例した時間で見つけることができる。なお、(1)、(3)、(4)では、子節点の値が必ず親節点よりも大きな値をとっているが、そのような木はヒープとよばれる。

問 29 正解【1】

ハッシュでは、ハッシュ関数が定数時間で計算可能である。ハッシュテーブルに登録されている要素数が  $n$  のとき、最悪の場合、要素の探索が  $O(n)$  にかかることがある。

問 30 正解【2】

リレーショナルデータベース(RDB)は表でデータを表現する。選択肢 2 は木構造なので、リレーショナルデータベースではなく階層型モデルのデータベースになる。

問 31 正解【4】

リレーショナルデータベース (RDB) は関係 (リレーション) でつなぎ、XML はタグでつなぐので、(a)と(b)の時点で選択肢 2 か 4 が候補。表の間の連携を取るのは、XML ではなく RDB なので、(c)が無くても選択肢 4 が正解と分かる。(c)は、検索を高速にするものなので、インデックスである。

問 32 正解【1】

ウェブを記述する言語で最も用いられているものは HTML である。

問 33 正解【2】

平均値は、 $(3+4+3+6+5) / 5 = 4.2$  である。

中央値は、整列データ (3、3、4、5、6) における中央の値である 4 である。

最頻値は 3 が 2 回、それ以外の数字が 1 回ずつ出現しているので、3 となる。

問 34 正解【1】

期待値演算子は線形演算子なので、選択肢 1 が成り立つ。詳しく解説するなら、期待値演算子の定義より

$$E[X] = \int x \cdot p(x)dx, \quad E[Y] = \int y \cdot p(y)dy,$$

$$E[X - Y] = \int (x - y) \cdot p(x, y)dxdy$$

がいえる。最後の等式は

$$E[X - Y] = \int (x - y) \cdot p(x, y)dxdy = \int x \cdot p(x, y)dxdy - \int y \cdot p(x, y)dxdy$$

と変形できるが、2つの項それぞれに、確率の話の公式

$$p(x) = \int p(x, y)dy, \quad p(y) = \int p(x, y)dx$$

を適用すると、

$$E[X - Y] = \int x \cdot p(x)dx - \int y \cdot p(y)dy = E[X] - E[Y]$$

を得る。

問 35 正解【1】

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は「両側検定」を行い、平均がある値 $\mu_0$ より大きいか検査する場合、もしくは小さいか検査する場合は「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、上回れば受容される。

問 36 正解【4】

選択肢 1-3: 母平均、母分散がどうであれ、最尤推定値は収集した標本に依存する。よって、標本を増やせば、平均パラメータ  $m$  や分散パラメータ  $v$  の最尤推定値は増えることも減ることもある。

選択肢 4: 標本を $x_1, \dots, x_n$ に対し、分散パラメータ  $v$  の最尤推定値は

$$v = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2, \quad m = \frac{1}{n} \sum_{i=1}^n x_i$$

で与えられる。 $n$  個の項は非負なので、分散の最尤推定値は必ず非負である。

問 37 全員正解としました。

当初はベイズの定理に関する語句の組み合わせを含んだ問題でしたが、問題文の推敲課程で違いを生じる設問を削除してしまったため、結果的に正解をひとつに絞れない選択肢となっていました。そのためこの問題は全員正解とします。

問 38 正解【4】

感度と特異度は、それぞれ  $TP/(TP + FN)$  および  $TN/(TN + FP)$  で定義される。この定義より、TP, FP, FN, TN は非負の整数をとることから、感度もしくは特異度は常に 1 以下の値になる。

問 39 正解【3】

識別関数の閾値を 0 とすると、 $f(x;w,b) > 0$  ならばそのデータは陽性と予測し、さもなければ陰性と予測される。 $x=(2,1)$  および  $x=(1,2)$  を識別関数に代入すると、識別関数の値は  $f(2,1;w,b)=1 > 0$ ,  $f(1,2;w,b)=-1 < 0$  となることから、 $x=(2,1)$  は陽性、 $x=(1,2)$  は陰性と予測される。

問 40 正解【3】

選択肢 1,4 は非階層的、選択肢 2 は階層的クラスタリングの方法であり、選択肢 3 はクラス分類手法である。

問 41 正解【1】

HapMap は元々ヒトのハプロタイプ [同一染色体上で連鎖する 1 塩基多型 (SNP) の並び] を集積した地図を構築する国際プロジェクトの名称で、そこで構築されたデータベースも HapMap と呼ばれている。

問 42 正解【4】

塩基配列のドットプロットでは、正方向と逆 (相補鎖) 方向の配列比較を行い、一致した領域がそれぞれ左上から右下、および左下から右上の斜線として表されている。2 つの配列に同一のパリンドローム配列が存在する場合、そこは正方向でも逆方向でも一致するので、X 字型のプロットが現れるはずである。

問 43 正解【3】

バルジループと内部ループは似ているが、バルジループは一方の鎖のみに塩基対を作らないループが存在する。RNA の 2 次構造予測の計算の際は、これらのループの種類ごとに異なるエネルギー関数が定義されている。

問 44 正解【4】

blastp はタンパク質配列をクエリとして、タンパク質配列データベースに対して検索を行う。

問 45 正解【3】

BWT は、接尾辞配列を作る要領で元の文字列のすべての接尾辞を辞書順にソートし、各接尾辞の直前の文字（先頭文字の直前は最後の文字とする）をとってつなげた文字列と定義される。たとえば元の文字列を **cgcgga** とすると、終端記号 **\$** を加えた **cgcgga\$** の変換後の文字列は **agg\$cc** となる（右図参照）。文字列の並べ替えの一種だが、この文字列から元の文字列を復元することが可能であり、その意味で可逆変換である [詳細は省くが、復元は、BWT 列と各接尾

接尾辞配列	BWT	接尾辞
6	a	\$
5	g	a\$
3	g	cga\$
1	\$	cgcgga\$
4	c	ga\$
2	c	gcga\$

辞の先頭文字の列 (=BWT 列をアルファベット順に並べ替えた文字列。ここでは **\$accgg**) との対応づけを繰り返すことによって行われる。図の矢印は文字列を後ろから復元する過程の一部を示している]。このとき、変換後の文字列では同じ文字が繰り返す傾向があることから、BWT は圧縮率を高めることが知られている。一方、BWT 列は、接尾辞配列と同様に、ゲノム配列などに対する効率的な索引付けの目的でも用いられる。その際は、復元のとくと同様の考え方で、検索文字列を後ろから順に照合するアルゴリズムが用いられるが、あらかじめ作成した FM-index という索引を用いることにより、長さ  $L$  の検索文字列に対して、ゲノムの大きさによらず  $O(L)$  の計算量で検索できることが知られている。これは、二分探索を用いた接尾辞配列による検索が、ゲノムの長さ  $M$  に依存して  $O(L \log M)$  であるのと比べて改善されている。

問 46 正解【4】

FASTA 形式は配列情報を格納する標準的なファイル形式だが、クオリティスコアは含まない。これを、配列に加えてクオリティスコアも格納できるようにしたのが FASTQ 形式である。SFF (Standard Flowgram Format)形式は元々 454 シークエンサの出力を格納するために開発された形式で、配列とクオリティスコアの情報を含んでいる。BAM 形式はシークエンサから出力された配列をリファレンス配列と照合してマッピングした結果を格納するための形式だが、元の配列とクオリティスコアの情報を保持できるようになっている。

問 47 正解【2】

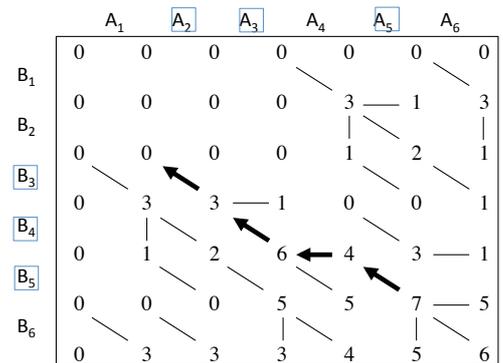
$k$ -mer 検索では、一般に  $k$  を大きくするほど偶然に一致する確率が小さくなる。例えば、塩基配列ですべての  $k$ -mer の出現頻度が等しいと仮定すると、 $k=5$  だと同じ文字列はおよそ 1kb ( $4^5=1024$ ) に 1 回出現するのに対し、 $k=10$  だと 1Mb に 1 回、 $k=15$  だと 1Gb に 1 回となり、偶然の一致による偽陽性の数が大幅に減る。これによって正確なアラインメントを計算する回数が減るので高速化が期待できる。一方、 $k$ -mer 検索ではミスマッチを考慮しないため、 $k$  が大きいほどミスマッチを多く含む配列を取りこぼす可能性が大きくなり、検出感度は低下する。

問 48 正解【1】

配列間の距離が、系統樹上の距離（枝の長さの和）で表されると考えて式を立てる。AB 間の距離は  $a+b$  で表されるので、 $a+b=30$ 。同様に AC 間の距離から  $a+c=40$ 、BC 間の距離から  $b+c=50$ 。連立方程式を解くと  $a=10$ 、 $b=20$ 、 $c=30$  と求められる。なお、配列が 4 本以上になると式の数が増えるので、連立方程式では解けず、最小二乗法などで枝の長さを決める必要が出てくる。

問 49 正解【3】

動的計画法で各点のスコアを計算した後で、最後にアラインメントを構築する過程にかんする問題である。スミス・ウォーターマン法では、局所的に最高スコアをとるアラインメントを求めるので、最高のスコアをとる点（ここでは 6 行 6 列目の 7）から始めて、動的計画法の過程で記録されたパスをたどって、スコア 0 の点に到達するまでのパスをとる（図の太線矢印）。左上に向かうパスは文字の照合、左や上に向かうパスはギャップに相当する。これらによってアラインメントを構成すると選択肢 3 が正解とわかる。



問 50 正解【2】

2つの遺伝子機能 X, Y において、X is\_a Y の関係があるとき、Y は X の上位概念（より一般的な概念）であり、X の機能を持つ遺伝子は、つねに Y の機能を持つということもできる（機能を持つ遺伝子の集合としてみると、Y は X を包含している）。そこで、“mitotic cell cycle process”と“nuclear DNA replication”をともに包含する上位概念を探せばよい。これはグラフを上方向にたどってどちらの点からも到達できる点に相当するが、ここでは“cellular process”と“cell cycle process”の2つがある。このうち“cell cycle process”の方が下位なので、より狭い概念であり、共通機能をよりの確に表すといえる。

問 51 正解【4】

次世代シーケンサによる解析の多くは、既知ゲノム配列をリファレンスとして、それへのマッピングを行うのが基本であるが、新規のゲノム配列を決定する際など、リファレンス配列がない場合は、リード配列どうしを照合してつなげる *de novo* アセンブリが必要となる。

問 52 正解【3】

重み行列に従って各位置のスコアを加算してスコアを計算する。CGTG は  $8+9-8-29=-20$ 、GTGT は  $-8-17+15-34=-44$ 、TGTA は  $4+9-8+11=16$ 、GTAG は  $-8-17-35-29=-89$  で、TGTA が最高スコアとなる。たいていの場合、真のモチーフ部位は正、それ以外は負のスコアとなるので、スコアの正負に注目すれば、厳密な計算を省略して比較的早く正解にたどり着ける。

問 53 正解【4】

溶媒分子を球で近似した時に、「溶媒接触表面」は溶質分子に接触した溶媒球の中心、また、「コノリー表面」は溶媒球の表面がなぞる面として定義され、それぞれ(A)図と(B)図の太線に相当する。二つの溶質分子が解離すると、溶質分子どうしの結合により溶媒分子が入り込めなかった部位に溶媒分子が結合できるようになるため、溶媒接触面積は増加する。以上のことから、選択肢 4 が正解である。

問 54 正解【4】

アミノ酸残基どうしが密に接している領域が 2 つ存在し、その接し方のパターンが類似しているため、このタンパク質は類似した構造を持つ2つのドメインからなることがわかる。そのうちの、N 末側ドメインの N 末（1~10 残基付近）と C 末（70~80 残基付近）の関係を示す領域には、対角線に垂直な黒線が見られるため、この部位は平行  $\beta$  シートではなく、逆平行  $\beta$  シート構造を形成していることがわかる。よって、選択肢 4 が正解である。

問 55 正解【2】

グロビンフォールドは $\alpha$ ヘリックス、免疫グロブリン(Ig)フォールドは $\beta$ ストランド、TIM バレルフォールドとフェレドキシンフォールドは $\alpha$ ヘリックスと $\beta$ ストランドからなる。

問 56 正解【1】

PDBML フォーマットでは、その内容が XML フォーマットで記述されている。実データを表す文字列を”<”（開始タグ）と”>”（終了タグ）で挟み込んで記述することで、表示用ソフトウェアに対し、文書の構造、書式、リンク情報などを指定している。コンピュータによる処理には適しているが、人間にとっては読みにくい。

問 57 正解【4】

$\alpha$ ヘリックスと $\beta$ ストランドが交互に現れるタンパク質は、SCOP において $\alpha/\beta$ のクラスに分類され、(多くの)平行 $\beta$ シートを有するという特徴を持つ。図は $\alpha/\beta$ の代表的なフォールドの一つ TIM バレルフォールドである。樽状の円筒形が特徴である。

問 58 正解【2】

ドメイン構造に見られる階層性は、様々なドメインを分類整理するための一つの見方である。この階層性を利用した代表的な立体構造分類データベースには SCOP と CATH がある。問題中で図示したように、両者の階層は良く対応している。しかし、CATH におけるアーキテクチャーは二次構造の空間配置の類似性は考慮するが、アミノ酸配列上での出現順序は考慮しない点において、SCOP と異なる。

問 59 正解【4】

任意の 2 つのタンパク質のアミノ酸配列の一致度と立体構造の相違度についての研究から、配列一致度が 30%以上であれば、RMSD は数 Å 以下であることが多いことが示されている。ホモロジー検索法では検出できないほど配列同士の一致度が低下してもなお、互いの立体構造は類似している例が多く報告されている。従って、アミノ酸配列の一致度が 30%未満であっても、互いが相同でないとは断定するのは非常に困難である。逆に、立体構造同士の高い類似性を根拠に、両者が相同であることが推定された例もある。

問 60 正解【1】

アミノ酸配列間で有意な類似度を示す二つのタンパク質は、構造上の類似性を有することが多い。この経験則に基づき、類縁タンパク質構造から目的のタンパク質構造を予測する手法をホモロジーモデリング法という。有意な類似性を示すタンパク質が見つかった場合、ホモロジーモデリング法が最も精度の高い構造を予測する可能性が高い。

問 61 正解【3】

フォールド認識法は、ホモロジーモデリングと同様に、問い合わせ配列のフォールドと類似した鋳型構造を立体構造データベースから選択することで立体構造を予測する。鋳型構造が配列レベルでのホモロジー検索で容易に発見される場合がホモロジーモデリング、そうでない場合がフォールド認識とされる。フォールド認識法は、主に相同な既知タンパク質が存在せず、ホモロジーモデリングを適用できない場合に使用されることが多い。

問 62 正解【3】

MRI（核磁気共鳴画像法）は、核磁気共鳴現象を応用し、生体内の情報を画像化する方法である。

問 63 正解【4】

$\alpha$ ヘリックスはアミノ基と 4 残基はなれたカルボニル基が水素結合を繰り返し形成するため、近傍で各残基がコンタクトしている。そのためコンタクトマップにおいて、左上から右下にかけての対角線上にコンタクトしているマス目が出現する。逆平行 $\beta$ シートは、右上から左下方向に帯状にコンタクトしているマス目が見られる特徴を持つ。平行 $\beta$ シートは、左上から右下方向に帯状にコンタクトしているマス目が見られる特徴を持つ。

問 64 正解【3】

タンパク質の主鎖の二面角  $\varphi$  と  $\psi$  と二次構造の  $\alpha$ ヘリックス・ $\beta$ シート構造の関係を理解することが求められる。 $\beta$ シートは $\alpha$ ヘリックスと比べて平面構造に近いので、 $\varphi$ 、 $\psi$ ともに 180 度に近い。

問 65 正解【2】

連鎖不平衡は、2つの対立遺伝子の遺伝子座が同一染色体上の近傍に存在したり、機能的に高度な依存関係が存在したり、交配のランダム性が不十分であったりなどが原因で生じる現象である。よって選択肢 2 がもっとも不適切であることがわかる。

問 66 正解【3】

遺伝的多型には 1 塩基多型、反復配列多型など多くの種類があるが、いずれも異なる複数の遺伝子型が集団内に世代を超えて共存しているものと定義されるため、遺伝する可能性が無視できる体細胞変異は除外される。よって選択肢 3 がもっとも不適切である。

問 67 正解【1】

子は両親からそれぞれの半数染色体を受け継いでいるはずである。バンドパターンから父親(♂)はホモ接合、母親(♀)はヘテロ接合であるので、染色体の転座などが無いとすると、子には父親のバンドと母親のバンドのうち 1 本が認められるはずであり、これを満たすのは選択肢 1 の子である。

問 68 正解【3】

分子系統樹は種分化や遺伝子重複による現存遺伝子の生成過程を木構造で表すことで、それらの類縁性を表現する。無根系統樹の根の位置を特定するには、解析対象の遺伝子のグループに含まれないことが明らかな遺伝子(外群)を系統樹に加え、外群の枝が接続される位置を根とする方法がある。したがって選択肢 3 の組み合わせが最も適切である。

問 69 正解【4】

双方向ベストヒットは 2 つの遺伝子が、どちらをクエリにしても他方がもっとも類似する遺伝子として見つかることを意味し、オーソログ(直系)関係を推定する基準の一つである。ただしこれは、遺伝子重複によるパラログ(傍系)の発生が比較的低頻度に起こり、なおかつ種分化よりも過去に起こっていることを暗に仮定しているため、遺伝子重複が頻繁に起こっている遺伝子群への適用は好ましくない。よって選択肢 4 の組み合わせがもっとも適切である。

問 70 正解【2】

系統樹推定法には、進化距離に基づくもの(UPGMA 法、NJ 法など)と塩基配列などの文字置換に基づくもの(MP 法、ML 法など)がある。また方法によって、進化速度の一定性(UPGMA 法など)や総置換数の最少性(MP 法など)などを仮定するが、その仮定が成立しない場合は推定の精度は低下する。オーソログとパラログが混在する場合、それらの進化速度は一定でない可能性がある。よって選択肢 2 の組み合わせがもっとも適当である。

問 71 正解【1】

問題のノードの下に連なる現存遺伝子の塩基だけを調べることで解決する。問題のノードが C であると仮定すると直下右の枝で C→G の 1 回、G であると仮定すると直下左の枝で G→C の 1 回、T であると仮定すると、直下右の枝で T→G かつ直下左で T→C の 2 回が最少置換数になるので、選択肢 1 の(A, C, G, T) = (2, 1, 1, 2)が正解である。

問 72 正解【2】

外群の接続する枝がルートであるので選択肢 1 は正しい。 $\alpha$ -および $\beta$ -グロビン遺伝子の出現は遺伝子重複によるが、ニワトリ、マウス、ヒトの遺伝子がそれぞれ 2 セット存在することから、重複が 1 回であること、および $\alpha$ -、 $\beta$ -グロビン遺伝子群のいずれもニワトリを含むことから、その遺伝子重複は哺乳類の出現以前に起こったことがわかる。よって選択肢 3 と 4 は正しい。ノード 2 は鳥類(ニワトリ)とその他の脊椎動物、ノード 5 は哺乳類内部での霊長類(ヒト)とげっ歯類(マウス) の分岐にそれぞれ相当するので、選択肢 2 の内容は不適切である。

問 73 正解【1】

次世代シーケンサから得られた配列は、精度が低い配列を多く含んでいる。従って、クオリティチェック (QC) を行い、パスした配列のみを以降の解析に用いるので不適切な選択肢は 1 となる。残りの選択肢は全て正しい。

問 74 正解【4】

ヒトゲノムの中には繰り返し配列が複数存在する。例えば Alu (約 350bp) と呼ばれる配列は、ヒトゲノムに数十万存在している。従って、短いリードをどれだけ積み上げてもこのような配列は超える(繰り返し配列の開始より上流と終了より下流を一つのリードとして読むことで上流と下流を一意につなぐ)ことができず、染色体ごとに 1 本の配列にすることはできず、4 は誤りである。

問 75 正解【4】

オーミックス解析は、大まかにシーケンサや DNA マイクロアレイといった核酸の性質を利用するゲノム・トランスクリプトーム解析と、物質の質量や化学構造を利用するプロテオーム・メタボローム解析に分類できる。選択肢 3 の質量分析装置は、代謝産物等のメタボローム解析に使われる。選択肢 4 のキャピラリーシーケンサは、DNA 配列を決定する機器であり、メタボローム解析には利用されない。

問 76 正解【4】

ヌクレオソーム構造を見るためには、例えば、ヒストンに巻き付いていない領域を消化する酵素を用いた MNase-seq や DNase-seq が用いられる。バイサルファイトシーケンシング法は、非メチル化シトシンがウラシルに変わることを利用して、DNA のメチル化状態を調べるために用いられる手法である。したがって選択肢 4 の記述は不適切である。

問 77 正解【3】

サポートベクトルマシンでは正例、負例のラベルが付与された教師データないし一部にラベルが付与された半教師データから判別超平面の学習が行われる。このため、教師なし学習とする選択肢 3 は誤りである。

問 78 正解【2】

ベイジアンネットワークでは有向非循環グラフ (Directed Acyclic Graph) により確率変数間の条件付き独立性が記述される。有向非循環グラフは定義上、有向閉路を含まない構造であるため、選択肢 2 の記述は不適切である。

問 79 正解【1】

グラフィカルモデルでは、設問の図にあるように 3 変数  $X_1$ 、 $X_2$ 、 $X_3$  について変数  $X_1$  と  $X_2$  から変数  $X_3$  へ有向辺が引かれた  $v$  構造となる時、変数  $X_3$  が未観測の場合は変数  $X_1$  と  $X_2$  は確率的に独立であるが、変数  $X_3$  が観測された状態では、確率的に独立ではない。このため、選択肢 1 の記述は不適切である。また、前記状態において有向の辺の向きが 1 つでも逆になった場合、変数  $X_3$  が未観測であれば変数  $X_1$  と  $X_2$  は確率的に独立であり、変数  $X_3$  が観測されていれば変数  $X_1$  と  $X_2$  は確率的に独立ではない。

問 80 正解【1】

設問中にある  $y(t)$  の一般解は  $\lambda_1$  と  $\lambda_2$  が共に正の時のみ第二項が 0 に収束し定常状態となる。また、この時の一般解の収束値は  $A^{-1}\mathbf{b}$  である。このため、選択肢 1 のみ正しく、他の選択肢は誤りである。

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。