

平成 21 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 解答【1】

コスミドベクターは、 $\lambda$ ファージの性質を用いて、プラスミドベクターよりも長い DNA 断片 (40kbp 程度) を扱うために用いられる。

問 2 解答【4】

外部の物質を分解してエネルギーを得るのは異化。同化はエネルギーを使って生体高分子を合成する作用。

問 3 解答【2】

生体膜中のコレステロール含量は膜の流動性に影響するが、コレステロールはリン脂質ではない。

問 4 解答【1】

抗原抗体反応を伴わない自然免疫は、かつては原始的な動物だけがもつ免疫システムと考えられていたが、現在ではヒトを含む多くの生物において、初期防衛機構として重要な役割を果たしていることが知られている。

問 5 解答【4】

10,000 塩基対の DNA には 20,000 個の塩基が含まれ、シャルガフ則によりその中で A と T、G と C の含量はそれぞれ等しい。従って、A は 30% 含まれることになる。

問 6 解答【1】

小胞体はタンパク質の生合成にかかわっているが、ATP の合成はミトコンドリアで行われている。

問 7 解答【3】

ヒトなどの減数分裂においては、染色体数は体細胞の  $1/2$  になる。

問 8 解答【4】

DNA の二重らせん構造は塩基間の水素結合により形成される。

問 9 解答【2】

DNA 複製は細胞周期の S 期 (S は synthesis から) に起こる。

問 10 解答【2】

イントロンは mRNA 前駆体の一部として、エキソン同様に転写されるが、mRNA が成熟する過程で切り取られる (RNA スプライシング)。

問 11 解答【1】

典型的な開始コドンは **ATG** で、メチオニンをコードする(コドン表を全部覚える必要はないかもしれないが、メチオニンと終止コドンぐらいは覚えておこう)。

問 12 解答【3】

リボソームは多数のタンパク質と **RNA** からなる複合体である。

問 13 解答【4】

核内受容体は、ステロイドホルモンなどのリガンドと結合した後、核内に移行し、**DNA** の特定領域と直接結合して作用するので、キナーゼ活性は必要ない。

問 14 解答【3】

ゲノム解読では、遺伝子間領域を含む全ゲノム塩基配列を決定しようとする。タンパク質をコードする領域を推定するために、トランスクリプトーム解析が併用されることが多い。

問 15 解答【3】

トランスクリプトーム解析の材料となるのは、転写物(**mRNA**)から作られた **cDNA** である。

問 16 解答【1】

二次元電気泳動法は、プロテオーム解析の基本的な手法の一つで、細胞内に存在するタンパク質をできるだけ区別するために用いられる。

問 17 解答【1】

ユビキチン化は、タンパク質に対する分解反応の目印として起こる反応である。

問 18 解答【4】

ウエスタンブロッティング法は、抗体を使って、(**mRNA** ではなく)細胞内の特定のタンパク質の存在を検出する方法である。

問 19 解答【2】

転写因子は **DNA** 結合タンパク質であり、**DNA** の切断は行わない。いわゆる核内受容体は転写因子として作用する。

問 20 解答【3】

X線結晶構造解析では、直接的にはX線が試料を構成する電子によって散乱され、結晶構造によって決まる条件により強め合う現象を観察している。従って、試料の形は電子の分布として表現される。

問 21 解答【3】

国際コンソーシアムによるヒトゲノム計画では、(セセラ社でも)当時一般的だったサンガー法が用いられた。パイロシーケンシング法は、いわゆる次世代シーケンサで用いられている方法の一つである。

問 22 解答【1】

連続する命令群に注目し、命令単位以下まで細分化するのは、演算パイプライン方式。マルチスレッドでは、アプリケーションを同時実行可能な短い処理の文脈(スレッド)に分割するが、通常1つのスレッドには多数の命令を含む。

問 23 解答【2】

インタープリタ型言語の例は Perl や BASIC。コンパイラ型言語の例は C, FORTRAN 等。

問 24 解答【2】

各桁ごとの論理演算において、0 との OR 演算では元のビットの値は不変。1 との OR 演算では結果は必ず 1 になる。2桁目と3桁目が元々一致しているのは、選択肢2のみ。

問 25 解答【3】

IP アドレスの枯渇は世界的に危急の問題。IPv6 と呼ばれる新しい通信プロトコルが普及すれば一気に解決するが、その普及に時間がかかっている。

問 26 解答【2】【3】

(お詫び:この問題には正解が複数存在しました。選択肢2と3を正解とします)

選択肢2は、根からすべての葉が出ている木を考えれば、葉の数は  $n-1$  となり、誤り。選択肢3も、一カ所だけが深い階層を持つ木を考えれば枝数は  $n-1$  となるため、誤り。

選択肢4は、根の2つの子節点の子孫節点の数が  $n/2$  より大きければ、全体の節点の数が  $n$  より大きくなってしまうため、どちらかの子孫節点の数は  $n/2$  以下である。

問 27 解答【4】

クイックソートは、ピボットとして取る値が常に最小値あるいは最大値であったような場合に最悪  $O(n^2)$  の時間がかかる。 $O(n \log n)$  はあくまでも平均的な状況での評価である点に注意が必要。それに対し、ヒープソート、マージソートは最悪でも  $O(n \log n)$  で計算可能。

問 28 解答【1】

選択肢3, 4で説明されているように、ハッシュ法では値の衝突が起こりえる。

$n$  個のデータのほとんどが同じハッシュ値になるという極端なケースを考えれば、常に定数時間で処理することはできず、選択肢1は誤り。ただし、平均的にはランダムな入力に対して定数時間で探索が可能である。

問 29 解答【3】

入力数列 10 の後、さらに 1 が1回続く場合、状態は S0 となる。よって選択肢3は誤り。

問 30 解答【2】

結合(join)演算はリレーション間に施す演算の中でもっとも時間を要する。

結合演算を1命令で実行できるような一般的な CPU は存在しない。

問 31 解答【4】

リレーショナルデータベースとオブジェクトデータベースを組み合わせた製品が広く利用されている。

問 32 解答【4】

この SQL 文が実行されると、サブタイプ欄が H1N1、かつ、年号が 1976 年以降 2005 年未満の項目が選ばれて、アクセッション欄が列挙される。

問 33 解答【3】

サイコロの目の出方は、過去の出方とは独立なので選択肢3は明らかに誤り。

選択肢2で、1回サイコロをふって出る目の期待値は 1 から 6 の平均で 3.5 である。

問 34 解答【1】

分散  $0.04g^2$  で標準偏差は  $0.2g$ 。誤差の絶対値が標準偏差の 2 倍の  $0.4g$  となる確率を調べるには、表中の 2.0 の欄(3行1列)を参照し 0.0228 を得る。ただし両側なのでその2倍。

問 35 解答【1】

$X^2$  の期待値  $E(X^2)$  が 5 で、 $X$  の平均  $E(X)$  が 1 ならば、 $X$  の分散は  $E(X^2) - E(X)^2 = 4$ 。  
標準偏差はこの平方根で 2 となる。したがって選択肢 1 は誤り。

問 36 解答【2】

抽出した標本の平均値(標本平均)は、多数の標本をとれば母集団の平均値(母平均)に徐々に近づくが、母分散が 0 の場合(全データの値が等しい場合)を除き、一般には誤差が生じる。よって、  
選択肢 2 は誤り。

問 37 解答【1】

線形カーネルでは識別面は(超)平面になるので、選択肢 1 は誤り。  
選択肢 3 で、 $p=1$ ,  $c=0$  の場合は、線形となる。

問 38 解答【1】

3層パーセプトロンは学習機械の一種であり、二分探索法とは関係がない。

問 39 解答【3】

近傍数  $k=1$  の場合には  $X_5$  のみ、近傍数  $k=3$  の場合には  $X_5, X_4, X_3$  の 3 データが参照される。前者では陽性、後者では陰性の判定となる。

問 40 解答【1】

出力確率行列  $B$  の定義により、文字  $A$  は状態  $Q2$  のみで出力され、文字  $T$  は状態  $Q3$  のみで出力されることがわかる。文字  $G$  および  $C$  は  $Q1$  と  $Q3$  のどちらからも出力可能だが、観測系列  $GCAT$  については、先頭の  $G$  と  $C$  は  $Q1$  から出力される場合しかあり得ない。

問 41 解答【4】

「FT SIGNAL 1 22」の記述は N 末端のシグナルペプチドを示しており、C 末端は関係ない。  
また、膜蛋白質ではシグナルペプチドが切断されない場合が多い。

問 42 解答【2】

**blastx** は DNA のクエリ配列を仮想翻訳(可能な 6 つの読み枠を仮定して標準的コドン表を用いた翻訳)し、データベースに収納されているアミノ酸配列に対して類似配列探索を行う。**BLAST** はよく使われるプログラム群であり、**blastn**, **blastp** などのプログラム名を覚えると便利。

問 43 解答【4】

この問題は、モチーフに現れる C の距離制約に注目すると解きやすい。

問 44 解答【2】

アラインメント・スコアは、ギャップの数×ギャップ開始ペナルティに、列毎のスコア(ペナルティ)を足して計算される。

問 45 解答【3】

DNA 配列でいう「回文」とは、相補鎖の逆読みが自分自身と同じ文字列である。

問 46 解答【2】

RNA 二次構造のモチーフは主鎖と水素結合の形で定義されている。この問題の選択肢にある、代表的な RNA 二次構造モチーフの名称を覚えておこう。

問 47 解答【3】

DNA にも見られる、3つの水素結合を形成する C-G は最も強い。A-U(DNA の A-T に相当)は2つの水素結合を形成し、C-G より結合は弱い。RNA 独特の塩基対である G-U も2つの水素結合を形成する。

問 48 解答【1】

マイクロ RNA は近年注目されている。その基本的な特徴を覚えておこう。

問 49 解答【2】

局所アラインメントでは、配列同士のもっとも似ている部分だけに注目し、アラインメントを行う。進化的に離れている生物の相同な配列は、一部にしか類似性が残らないことが多いから、ホモロジー検索は普通局所アラインメントを使用する。ペアワイズアラインメントの計算量は  $O(L^2)$  である。

問 50 解答【4】

最適アラインメントは2本の配列が進化的に共通した祖先を持つと仮定し、分岐後、現在に至るまでの、もっとも確率の高い変異経歴を推定する。BLOSUM と PAM はタンパク質のアミノ酸置換行列であり、核酸の比較に使われない。PAM20 の「20」や PAM70 の「70」は、平均的にみた、100 残基当りに進化的に受容されるアミノ酸置換の数を意味する。従って、PAM は数字が大きいほど、進化的に遠く離れている配列に適している。BLOSUM65 などの「65」は、PAM と違って、数字が小さいほど進化的に遠く離れている配列に適している点に注意が必要。

問 51 解答【2】

動的計画法はバイオインフォマティクス以外にも情報処理に広く使われる一般的な手法である。動的計画法の使える問題の共通点は、独立に解いた部分問題の解を、部分問題間の依存性を考慮せずに、簡単に組み合わせられることである。

問 52 解答【2】

n本の配列アラインメントは動的計画法より $O(L^n)$ で厳密に計算できる。しかし、マルチプルアラインメント( $n \geq 3$ の場合)には、この計算量が莫大になることが多いため、通常はより高速な近似的手法が使われる。

問 53 解答【3】

ペアワイズアラインメントのスコアは配列とアラインメントパラメータで決まるが、e-値はデータベースの大きさにも依存する。これは、データベースが大きければ大きいほど、偶然クエリと似ている配列が含まれる確率が上がる、という効果を反映している。

問 54 解答【3】

分泌されるタンパク質の多くは、N末端付近にシグナルペプチドを持ち、ミトコンドリアと葉緑体に局在する多くのタンパク質もN末端付近の局在シグナルにより細胞内局在が決まる。

問 55 解答【3】

ドットマトリックス図で配列間の類似部分を可視化すると、同じ方向で読んで類似する部分是对角線、片方の配列の相補鎖を逆読みして類似する部分は反対角線、と並行する線分として表される。

問 56 解答【2】

AMBER, BLAST, HMMRED はそれぞれ、分子動力学シミュレーション、置換行列を利用した配列類似性検索、隠れマルコフモデルを利用した類似性検索法である。HMMRED は、最近ではHHPREDと呼ばれるようになっている。

問 57 解答【2】

3DCoffee は、配列のマルチプルアラインメントのプログラム。立体構造を考慮するので3D がついている。

問 58 解答【3】

左巻きヘリックスは側鎖の無いグリシンなど特殊な残基しか許容できない二面角をとるので、通常のタンパク質ではほとんど観測されない。

問 59 解答【4】

ATOM 行の(-0.603, 65.642, 77.183)が83番のバリンのN原子のxyz座標であり、その後ろに続く1.0が占有度、23.27が温度因子の値である。なお、ポリペプチド鎖が8本含まれることは、COMPND 行のCHAINの記述を見るとわかる。



問 60 解答【3】

Critical Assessment of ... という名前で行われている手法の国際的評価会に関する問題。構造予測(... Structure Prediction)の略で CASP が正解。タンパク質間相互作用の予測評価会である、CAPRI (... PRediction of Interactions)とマイクロアレイデータ解析法の評価会 CAMDA(... Microarray Data Analysis)とセットにして覚えておくと良い。ComFA は、Comparative Molecular Field Analysis の略で、分子の立体構造を考慮した構造活性相関 (QSAR) の解析手法の1種であるので関係ない。

問 61 解答【4】

「側鎖がなく自由度の高いグリシン」と「側鎖が主鎖の窒素に結合していて  $\phi$  が 60 度前後に固定されているプロリン」の 2 種類は他の 18 種類の標準アミノ酸とは大きく異なるラマチャンドランプロットを示す。

問 62 解答【4】

ほとんどのフォールドでは「1フォールド=1スーパーファミリー」である。多くのスーパーファミリーを含むフォールドの種類は限られており、スーパーフォールドと呼ばれているので、4が明らかな間違いである。

なお、SCOP では Class の階層で、2次構造の含量以外につながりの順番も考慮 ( $\alpha + \beta$  と  $\alpha / \beta$  を区別)しているのので、一見3も間違いに見えるが、2次構造の含量が一番主要な要素であることに違いはないので間違いではない。

問 63 解答【2】

X 線結晶構造解析の全体としての精度は、解像度と  $R_{\text{free}}$  値が良い目安となる。それぞれ値が小さいほど良い(選択肢1, 3)。各原子位置の精度を評価するには温度因子が便利である。温度因子は原子毎に計算され、座標を平均値と考えた時の標準偏差に対応する原子の位置のゆらぎで、小さい方が座標位置は正確である(選択肢2)。構造解析では最終的な位相のみが重要であり、どのようにして初期位相を決めたかは関係ない(選択肢4)。

問 64 解答【2】

近親交配がおこれば、同じ遺伝子型を持つ子供が生まれやすいので、ヘテロ接合体ではなく、ホモ接合体の割合が高くなる。

問 65 解答【4】

アジア系サンプルは親子トリオで取得されていない。

問 66 解答【1】

連鎖不平衡は、生物集団に対して定義されるもので、家系に対して定義されるものではない。

問 67 解答【2】

テロメアは染色体末端の構造であり、多型の呼び名ではない。

問 68 解答【4】

ウイルス表面抗原において、宿主の抗体の攻撃から逃れるためのアミノ酸置換の加速は、正の選択によって生じるもので、中立な変異ではない。

問 69 解答【4】

系統樹では、生物の並ぶ順番自体は意味を持たず枝の分岐パターンが系統関係を示している。系統樹をさかのぼったとき、ヒトとワニの枝が合流する節（ノード）は、ヒトとニワトリの枝が合流する節と同じである。

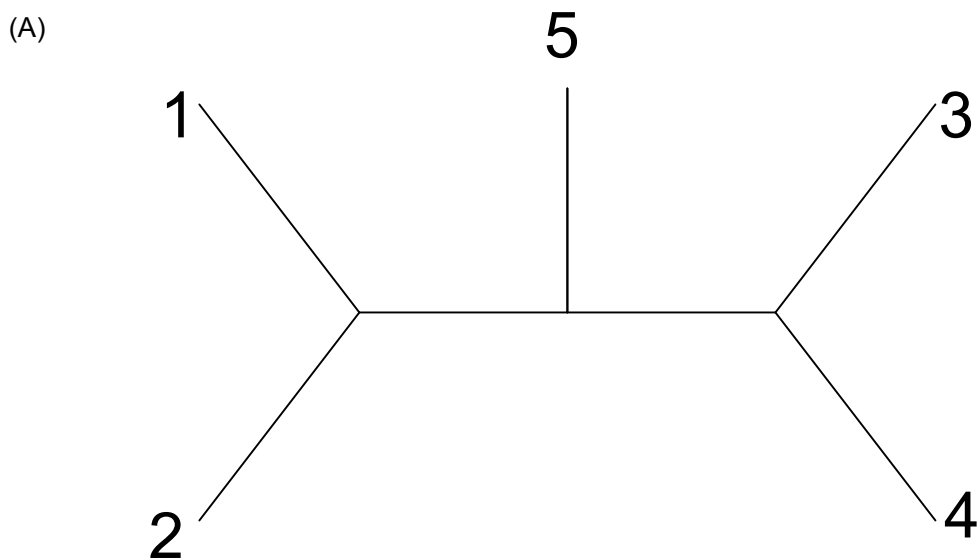
問 70 解答【1】

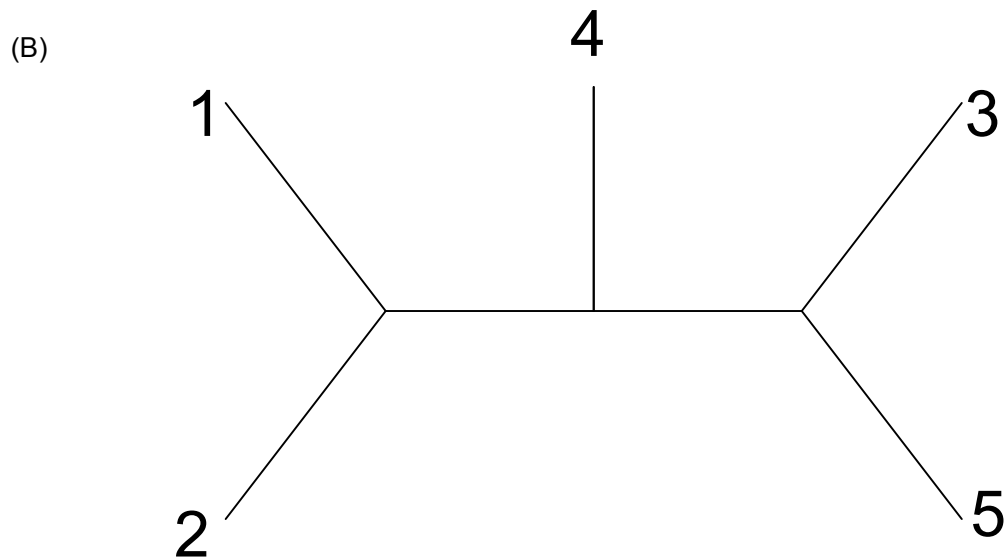
[イ] 置換が生じなかった割合なので、1から相違度  $K$  を引いたものになる。

[ロ] [イ] =  $e^{-k}$  の両辺の対数を取り、変形してやれば得ることができる。

問 71 解答【4】

説明に従い系統樹（無根）を構築してみると、問題分のもと選択枝 1,2,3 ではいずれもトポロジー (A) が得られる。選択枝 4 はトポロジー (B) の系統樹を表す。





問 72 解答【3】

種間で遺伝子の並びが互いによく似た領域をシntenニック領域と呼ぶ。

問 73 解答【2】

スパイクインコントロール法では、濃度既知のコントロールRNAをサンプルのRNA溶液に加えてから標識反応を行う。データ補正のほか、実験のトラブルシューティングにも利用される。

問 74 解答【2】

タイリングアレイとは、決定されたゲノム配列に基づいて稠密に設計したプローブを搭載したものである。

問 75 解答【1】

1サイクルごとにPCR産物は2倍に増えるので、一定量に達するのにサイクル数が2回少なくて済むということは、含有量が2の自乗、つまり4倍多く含まれていたことを示す。

問 76 解答【3】

BLASTPはアミノ酸配列の相同性検索に、DALIはタンパク質立体構造の比較に、それぞれ用いられるプログラムである。

問 77 解答【4】

mRNA のレベルとタンパク質のレベルが相関しないのは、mRNA 合成以降のステップに原因があるためと考えられるので、mRNA の合成速度には関係がない。

問 78 解答【4】

基質濃度がミカエリス定数に等しいと反応速度は最大値の  $1/2$  になるが、ミカエリス定数の2倍の濃度の基質があっても反応速度が最大にならない。 $S = 2K$  を式に代入すれば明らか。

問 79 解答【1】

$B \rightarrow G$  の流束は一定なので、 $E \rightarrow D$  の流束が最小になる場合(=  $C \rightarrow E$  の流束が全てHに向かう)と最大になる場合(=  $C \rightarrow E$  の流束が全てDに向かう)を考えればよい。

問 80 解答【3】

遺伝子  $\alpha$  が発現するのは、Xが発現している4つの区画のうち、YもZも発現していない区画である。よって、上段左端の区画のみとなる。