

平成 23 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】

水分子中の酸素原子が他の水分子の水素原子と行う主な相互作用は水素結合である。これはファンデルワールス結合よりもずっと強い。

問 2 正解【4】

サイクリック AMP は ATP から合成されるが、生体内では二次メッセンジャーの一種として、シグナル伝達に用いられている分子である。

問 3 正解【3】

イは小胞体、オはエンドソームなどの小胞を表す。

問 4 正解【1】

いわゆる原核生物(真正細菌と古細菌)の細胞はミトコンドリアを持たない。

問 5 正解【1】

M 期、S 期の名称はそれぞれ **mitotic phase, synthesis phase** に由来。テロメアは染色体の両端にあつて、染色体の末端を保護している。休止期(静止期)は **G0** と略記される状態。

問 6 正解【4】

ゲノムが複数の染色体に分割して担われている生物では、たとえ交叉が起こらなくても、父母由来の染色体の選ばれ方により、遺伝子が入り交じることが可能である。

問 7 正解【2】

ホメオティック遺伝子は複数の染色体上の近接した領域にクラスター上に存在することが多い。ヒトなどではそのクラスターがさらに複数存在している。

問 8 正解【2】

二次(セカンド)メッセンジャーとは、シグナル伝達において、ホルモン等の情報伝達物質が細胞外で受容体に結合したという情報が細胞内部に伝えられるときに用いられる物質のことで、mRNA とは関係がない。

問 9 正解【1】

フェリチンは水溶性の鉄貯蔵タンパク質で、血液検査などで鉄欠乏性貧血などを調べるのに用いられる。

問 10 正解【3】

ユビキチンは、いろいろなタンパク質と共有結合すること(ユビキチン化)によって、そのタンパク質の分解を促す働きをする。

問 11 正解【2】

アデニン、グアニンはプリンとよばれる。

問 12 正解【1】

転写と違って、DNA 複製では両方の DNA 鎖が鋳型となる(リーディング鎖とラギング鎖)。

問 13 正解【1】

mRNA は 5'から 3'方向にむかって合成される。

問 14 正解【4】

RNA スプライシングは、mRNA 前駆体(プレ mRNA ともいう)からイントロンが取り除かれて、成熟 mRNA ができる反応。

問 15 正解【2】

N-グリコシル化は通常、細胞外にあるタンパク質に起こるので、核内受容体には不似合いな反応であり、転写制御との関係も現在のところ特に知られていない。

問 16 正解【3】

PCR 法では、大腸菌などの細胞を使うことなく、目的の DNA を増幅することができる。その反応には好熱菌由来の DNA ポリメラーゼが使われる。

問 17 正解【4】

ベクターは、自然界に存在するプラスミドなどの天然の DNA に人工的に改変を施したものがよく用いられる。

問 18 正解【2】

プロテオーム解析とは、細胞内に存在するタンパク質を網羅的な解析を指すので、DNA シークエンサではなく、質量分析計などが用いられる。

問 19 正解【4】

ノーザンハイブリダイゼーションとは、核酸の相補性を利用して、目的とする配列をもつ RNA を検出する方法である。

問 20 正解【3】

ChIP-Seq 法では標的とする特徴を特異的な抗体によって検出するので、たとえば転写因子が認識し、結合する領域を検出できるが、転写因子などのDNA結合タンパク質一般をコードする塩基配列を知ることはできない。

問 21 正解【1】

回路図より(a)も(b)も 1

問 22 正解【2】

TCP/IP では機器や通信回路の状況によって通信速度が異なる

問 23 正解【1】

ローカルエリアネットワークへの接続に、メールアドレスは必要ない

問 24 正解【3】

メモリに入り切らない大きなデータを読み込むと、仮想のメモリが主としてハードディスク上に用意される。これをメモリスワップと言う。ディスクはメモリに比べアクセス速度が遅いため、見かけ上、速度が遅くなる。

問 25 正解【3】

適切なものは(a)インタープリタ、(b)コンパイル

問 26 正解【4】

クイックソートは平均 $O(n \log n)$ 時間要するが、データの入力順が最悪の場合は $O(n^2)$ を要する。一方マージソートを用いれば、必ず $O(n \log n)$ で終了する。

問 27 正解【4】

長さ m の文字列が、長さ n の文字列に含まれているか否かの探索は、例えば Knuth-Morris-Pratt 法を用いると $O(n+m)$ である。必ず $O(n \log n)$ が必要であるとは言えない。

問 28 正解【1】

ヒープでは、子の値は親の値より必ず小さくなければいけない。それを満たすものは 1 だけである。

問 29 正解【2】

AABB*** の場合のみ異なり、スタックでは A、キューでは B がとり出される。

問 30 正解【2】

連結リストで k 番目の要素を発見するには、 k 回ポインタをたどる必要があるため、 $O(n)$ 時間を要する。

問 31 正解【3】

選択肢 1、2、4 は SQL の結果がメタンであるが、3 はエチレンである。

問 32 正解【3】

一般に、索引を作成すると、索引の分だけデータを保存する容量が必要になる。

問 33 正解【3】

$p=0.3$ の時、尤度は $0.3 \times 0.7 \times 0.7 = 0.147$ 。一方、 $p=0.5$ の時は、 $0.5^3 = 0.125$ 。よって、 $p=0.3$ の方が大きい。また、尤度が最大となるのは、 $1/3 = 0.33\ldots$ である。

問 34 正解【1】

統計的独立と言っているので、 $P(A|B) = P(A)$ 。よって、 $P(A) < P(A|B)$ は成立しない。

問 35 正解【2】

標準偏差の値は常に 0 以上。

問 36 正解【4】

ベイズ推定はパラメータを点推定するためのものではない。

問 37 正解【4】

選択肢 4 が正解。正規分布の内、平均 0、分散 1 のものを標準正規分布と呼ぶ。

問 38 正解【4】

問題の指示に従って計算すると、選択肢 4 だけクラスタの割り当ての変更が起こらないので、選択肢 4 が正解。

問 39 正解【2】

X に最も近い 3 点の標本番号は、近い順に 3、1、4。そして、それらのクラスラベルは B、A、B なので、B を X のクラスラベルの予測とする。

問 40 正解【3】

leave-one-out 法は、データのうち 1 個のみをテストに残す。

問 41 正解【3】

通常の計算とは逆に、動的計画法のスコア表から、一致、不一致、ギャップの各スコアを推測する。まず最適経路上のセルを順に見ていくと一致が+2 でギャップが-3 であることが分かる(ギャップについては、1 行目と 1 列目に注目してもよい)。次に 5 行 5 列目のスコアが 2 のセルに注目すると、ここにギャップで到達したとするとスコアが 0 になるはずなので、ここは左上のセルから不一致(G-T)で到達したはずである。よって、不一致のスコアを x とすると $*+x=2$ (式 1)と表せる。次に「*」のセルに注目すると、ここも不一致(T-G)なので、動的計画法の計算式から、 $*=\max(1-3, 1-3, 4+x)$ (式 2)と表せるが、 $*=-2$ とすると式 1 から $x=4$ となり式 2 と矛盾する(-2 は選択肢にもない)。従って $*=4+x$ となり、式 1 と合わせて計算すると $x=-1$ 、 $*=3$ (選択肢 3)となる。

問 42 正解【3】

条件付き確率の公式 $P(A|B) = P(A \cap B)/P(B)$ に当てはめて計算する。 $P(A^1) = 0.07 + 0.05 + 0.05 + 0.07 = 0.24$ ($P(A^1)$ は 1 行目の合計値)、 $P(A^1 \cap G^2) = 0.05$ (1 行 3 列目の値)から、 $P(G^2|A^1) = P(A^1 \cap G^2)/P(A^1) = 0.05/0.24$ となる。

問 43 正解【2】

累進法は、案内木に従って配列を順に併合していくことによりアラインメントを組み上げていくが、一度挿入したギャップの位置は修正しないため、配列数が少ない初期段階でのエラーがそのまま残るという問題がある。逐次改善法は配列全部を使ってアラインメントを評価し改善していくため、累進法の結果をさらに改善する目的でも使われる。

問 44 正解【1】

国際塩基配列データベースは、日本の DDBJ、米国の GenBank、欧州の ENA が相互に協調して運営されている。なお、ENA (European Nucleotide Archive) は古くからある EMBL-Bank に Trace Archive と Sequence Read Archive (SRA)を合わせたもの。ArrayExpress は遺伝子発現などの機能ゲノミクス関連のデータベース。

問 45 正解【4】

PSI-BLAST は、検索された配列から位置特異的スコア行列を作成して再検索を行うことによって相同性検索の精度を向上させる。ブートストラップ法は、ひとつのデータからのリサンプリングによって多数のサンプルを生成して統計的な評価を行う方法で、系統樹の分岐パターンの信頼性評価等に用いられている。

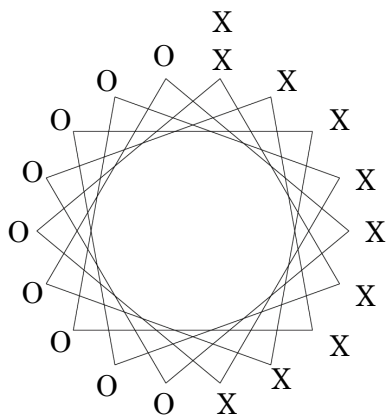
GenScan と Glimmer は、統計モデルを用いたコード領域予測プログラムで、それぞれ真核生物と原核生物のゲノムを対象とする。コード領域予測のもう一つの重要なアプローチは **cDNA** や既知遺伝子データベースへの相同性検索で、そこでは **BLAST** が有効。**AMBER** は分子動力学シミュレーションに用いられるプログラム。

ルールに従って、パターンと配列とを順に照合していけば解ける。なお、選択肢 **3** はパターンを含んではいるが、全長が一致するわけではないので不適。

アラインメントの左右にある配列位置を見ると、**Subject** に対して **Query** の方が 3 倍の大きさで変化しているので、クエリは核酸配列を翻訳したものであると分かる。核酸のクエリを翻訳してタンパク質データベースに照合するのは **TBLASTN** ではなく **BLASTX**。なお、選択肢 4 で、終止コドンは翻訳配列中で * によって表されている。

ドットプロットが左下から右上への線(／)であれば、右向き(→)の遺伝子は上向き(↑)に、左向き(←)の遺伝子は下向き(↓)にマップされ、左上から右下への線(＼)であればそれぞれ逆向きにマップされる。このルールに従って、ゲノム 1 上の遺伝子をゲノム 2 上にマップしていけばよい。

helical wheel 図の各頂点にアミノ酸残基を順に当てはめていき、OとXがそれぞれ異なる面に偏って存在するものを選べばよい。正解の選択肢 3 の場合、以下ようになる。



問 51 正解【2】

位置 2,3 のように全配列が同じ文字であるカラムは、エントロピーが 0 で、情報量は最大値の 2 ビットとなる ($1 \log 1 = 0 \log 0 = 0$ に注意)。また位置 5 も情報量が正確に計算できて、 $R(5) = 2 + \left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4}\right) = 2 - \frac{3}{2} = 0.5$ ビットとなる。従って答えは選択肢 2。別解として、頻度が高いものから小さいものへ度数を移動させて分布を均一化させると、一般にエントロピーは大きく、情報量は小さくなるので、たとえば位置 6 の分布(34,9,5,0)は位置 4 の分布(43,5,0,0)より情報量が小さい。これから情報量の大きさは $2=3>4>6>1,5$ となることが分かり、答えの選択肢 2 が導ける。

問 52 正解【4】

括弧の位置だけでなく、対応する括弧が塩基対と一致するように、括弧の向きにも注意して対応をとること。

問 53 正解【2】

普通、タンパク質分子には、安定的に水素結合した結合水が多数認められるので、選択肢 2 は不適切である。

問 54 正解【4】

触媒残基として多用されるアミノ酸残基は側鎖に解離基を持つので、選択肢 4 のイソロイシンがもっとも可能性が低い。

問 55 正解【4】

PDB, mmCIF, PDBML フォーマットは Protein Data Bank で使用される構造記述フォーマットであるが、選択肢 4 の FASTA フォーマットは配列用であり、構造(原子座標)を記述することができない。

問 56 正解【1】

由来は SOURCE 行に、解析手法は EXPDTA 行に、アミノ酸置換と欠失は REMARK 行に、2次構造は HELIX 行と SHEET 行に、ジスルフィド結合は SSBOND 行に記述されている。この場合、SHEET 行の2行目からこのシートが逆平行(-1)である事がわかるので、選択肢 1 は不適切である。

問 57 正解【1】

現在の水準の分子力学法では、分子軌道を考慮する場合はフラグメント分子軌道(FMO)法を用い、シュレーディンガー波動方程式を分子全体について解くことは行われない。よって選択肢 1 は不適切である。

問 58 正解【3】

ドメインは立体構造上まとまった球状領域であり、一次構造(配列)上で離れたペプチド領域から構成されるので、選択肢 3 は不適切である。

問 59 正解【1】

結合角(θ)はベクトル(化学結合 $\mathbf{a}$, $\mathbf{b}$)の内積 $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos \theta$ である事を使って求めるのが一般的である。よって選択肢 1 が正解である。

問 60 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、 $\{(1.0^2 + 0.0^2 + 1.0^2 + 3.0^2 + 2.0^2)/5\}^{1/2} = 1.7\text{\AA}$ (選択肢3)が正解である。

問 61 正解【4】

設問中の式から、この場合のチョウ・ファスマンパラメータは $\{300/(300+100+400)\}/\{3000/(3000+1500+4500)\} = 9/8$ (選択肢 4)が正解である。

問 62 正解【4】

鏡像異性体は、お互いに鏡に映した状態にある(すなわち重ね合わせることができない)分子で、アミノ酸においては主に $\text{C}\alpha$ 原子に結合した原子団が全て異なることに由来する。よって側鎖が水素であるグリシン(分子 $\mathbf{b}$ と $\mathbf{c}$)には鏡像異性が存在しないので、D-アラニン(分子 $\mathbf{a}$)と L-アラニン(分子 $\mathbf{d}$)の組み合わせ(選択肢 4)が正解である。

問 63 正解【2】

B 型 DNA は、図の様に塩基対の面がらせん軸とほぼ垂直をなす。塩基を数えると、図には 13 塩基対(26 塩基)が示されている。リン酸骨格をたどると、この DNA が右巻き(どちらに骨格をたどっても進行方向に見て右回りする)であることがわかる。矢印の指す溝の幅が他より広いことから、主溝(major groove)であることがわかる。よって選択肢 2 が不適切である。

問 64 正解【2】

リボンモデルで α ヘリックスはらせん、シート (β ストランド) は矢印として示されるので、これらの構造はそれぞれ、(a) α ヘアピン、(b) $\beta \alpha \beta$ 、(c) β ヘアピン、(d) β ヘリックスであることがわかる。よって選択肢 2 が正解である。

問 65 正解【4】

日本人を含め、トリオで取得していない民族も含まれる。

問 66 正解【3】

パブリケーション・バイアスはメタアナリシスに混入する大きなバイアスのうちの1つである

問 67 正解【4】

A 型頻度 = $0.1^2 + 2 \times 0.1 \times 0.4 = 0.09$, B 型頻度 = $0.5^2 + 2 \times 0.5 \times 0.4 = 0.65$, O 型頻度 = $2 \times 0.1 \times 0.5 = 0.1$, $0.4^2 = 0.16$ であることから、選択肢3が正解である。

問 68 正解【1】

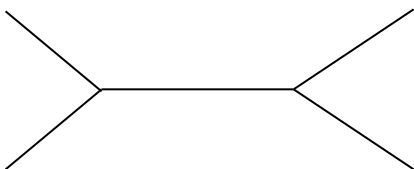
オーソログな配列でも種特異的に偽遺伝子になっていることがある。たとえば、ヒトを含む霊長類はビタミン C 合成能を欠いているが、これは L-グルノラクトンオキシダーゼ遺伝子が偽遺伝子化しているためである

問 69 正解【3】

動的計画法 (dynamic programming) は一般的な最適化アルゴリズムであるが、バイオインフォマティクスではアラインメントなどに使用されている。ただし、分子系統樹構築には直接的には使用されていないが、祖先配列推定に使用されている場合はある。

問 70 正解【1】

4 つの OTU からなる無根系統樹のトポロジーは次のようになる。



この系統樹5つの枝があり、そのいずれかに新たな OTU を付加することで5つの OTU からなる系統樹が得られる。4つの OTU からなる系統樹は、OUT の配置の異なる組み合わせから3種類あるので、5つの OTU からなる系統樹の総数は $3 \times 5 = 15$ である。

問 71 正解【2】

以下の連立方程式 $\alpha = d(A-O) + d(B-O)$, $\beta = d(A-C) = d(A-O) + d(C-O)$, $\gamma = d(B-C) = d(B-O) + d(C-O)$ より、 $\alpha + \beta - \gamma = 2d(A-O) + d(B-O) + d(C-O) - d(B-O) - d(C-O) = 2d(A-O) = 2d(O-A)$ が求められる。

問 72 正解【2】

オーバーラップ領域のある遺伝子の同義置換率は、フレームのずれから別の遺伝子の非同義置換による制約をうけることになるので、オーバーラップしていない領域の同義置換率よりも低くなる。

問 73 正解【2】

通常、SDS-PAGE 実行条件下ではタンパク質間相互作用は解消され、A と B は独立に泳動されるので、複合体形成の可能性は除外できる。

問 74 正解【3】

マイクロアレイ法では RNA 量を測定しているので、補正では合成速度を求めることはできない。

問 75 正解【4】

タンパク質では水素結合や疎水性相互作用により安定な立体構造を形成するが、RNA では tRNA など一部の例外を除き、安定な立体構造は形成しない。

問 76 正解【4】

質量分析装置はプロテオーム解析やメタボローム解析に用いられる。

問 77 正解【1】

モノクローナル抗体は特定のエピトープを認識する。エピトープが化学的に修飾されたり、スプライシングやプロテアーゼ消化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できなくなる。

問 78 正解【1】

システムノイズに影響を受けにくいモデルを、ロバストネスが高いモデルと呼ぶ。

問 79 正解【3】

ベイジアンネットワークでは非サイクルなネットワークを扱うので、回路を含む選択肢3が不適切である。

問 80 正解【2】

定常状態では左辺を0として $X(t) = \frac{a}{b}$ となるので、 $X(0)$ の値によらず a の増大と共に大きくなる。

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。