

**平成28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**  
**試験問題**

**注意事項**

1. 試験開始の合図があるまで、問題冊子を開いてはいけません。
2. 問題数は **80 問** です。すべての問題に解答してください。
3. 答案用紙への受験番号などの記入は、試験監督の指示があってから始めてください。
4. 試験時間は、**13 : 30 ~ 15 : 30 ( 2 時間 )** です。
5. 試験中にトイレに行きたくなったり気分が悪くなったりした場合は、手を挙げて試験監督に合図してください。
6. 退出可能時間は **14 : 30 ~ 15 : 20** です。  
途中で退出する場合には、手を挙げて試験監督に合図をし、解答用紙が回収されてから静かに退出してください。この場合、試験終了まで再入場はできません。
7. 試験問題の内容に関する質問にはお答えできません。文意通りに解釈してください。
8. 解答用紙（マークシート）の記入に当たっては、次の指示に従ってください。
  - (1) 問80までの解答欄を利用してください。問81以降への記入は、無効です。
  - (2) HB かB の黒鉛筆もしくはシャープペンシルを使用してください。
  - (3) **受験番号**は、受験票に印字されている通りに記入し、マークしてください。
  - (4) **氏名とフリガナ**は、受験票に印字されている通りに記入してください。
  - (5) **年月日欄**には、生年月日を記入してください（**年は、西暦で記入**）。
  - (6) 解答は、各解答記入欄に一つずつマークしてください。
  - (7) 訂正の場合は、あとが残らないように消しゴムできれいに消し、消しくずを残さないでください。
9. 試験終了後、この問題用紙は持ち帰ることができます。
10. **解答用紙（マークシート）**は、白紙であっても必ず提出してください。

問 1

二次元電気泳動を用いたタンパク質の分析に関する記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 分子ふるいの原理でタンパク質を分離する。
- 2 電荷の違いでタンパク質を分離する。
- 3 色素や抗体を用いて検出を行う。
- 4 密度の違いでタンパク質を分離する。

問 2

質量分析機を用いたタンパク質の分析に関する記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 安定同位体を用いてサンプル間でタンパク質の量を比較できる。
- 2 ペプチドの MS/MS 解析ではエドマン分解によってペプチドのアミノ酸配列を決定する。
- 3 ペプチドフラグメントの質量の分布パターンからタンパク質の同定ができる。
- 4 液体クロマトグラフィであらかじめペプチドを分離することで、より多くのタンパク質を同定することが可能になる。

問 3

ヒトの細胞に関する以下の記述のうちでもっとも適切なものを選択肢の中から一つ選べ。

- 1 ヒトの配偶子は一倍体である。
- 2 間期の細胞にはセントロメアとテロメアが 1 つずつ存在する。
- 3 核の無い細胞は存在しない。
- 4 体細胞は 46 倍体である。

問 4

全ゲノム配列決定に用いられるショットガンシーケンスに関する記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 リピート配列の決定は難しいことがある。
- 2 断片化された DNA をランダムに配列決定する方法である。
- 3 必ずしもベクターにクローニングする必要はない。
- 4 参照配列がなければ配列決定はできない。

問 5

減数分裂に関する記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 減数分裂に先立って DNA の複製が行われ、DNA 量は倍化する。
- 2 減数分裂では均等な細胞分裂が 3 回連続して起こる。
- 3 減数分裂では組換えが起こり、配偶子の遺伝的多様性が生ずる。
- 4 二倍体の生物が減数分裂を行うと一倍体の配偶子が生ずる。

問 6

DNA シーケンシングの原理について述べた以下の説明から、サンガー法の原理としてもっとも適切なものを選択肢の中から一つ選べ。

- 1 DNA ポリメラーゼがヌクレオチドを取り込む際に生じた微弱な pH の変化を半導体センサーで検出することで DNA ポリメラーゼによるヌクレオチドの取り込みの有無を判定する。
- 2 基板上に固定された DNA を鋳型に、可逆的ターミネータと呼ばれる蛍光標識ヌクレオチドを DNA ポリメラーゼによって 1 塩基ずつ取り込ませ、その蛍光を検出することで塩基配列を逐次決定する。取り込まれた蛍光標識はシグナル検出後に化学的に除去することが可能である。
- 3 ダイターミネータと呼ばれる蛍光標識ヌクレオチドが取り込まれた DNA フラグメントをポリマーが充填されたキャピラリーで分離することで塩基配列を決定する。
- 4 Zero-mode waveguides 法 (ZMW 法) を用いて DNA ポリメラーゼ近傍のごく限られた空間にのみ励起光を照射する。ヌクレオチドが DNA ポリメラーゼに取り込まれる際に蛍光標識が一時的にその場に留まることを利用して塩基配列を決定する。

問 7

mRNA の構造に関する次の記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 タンパク質の翻訳は mRNA の 5'末端側から 3'末端側に向かって進行する。
- 2 タンパク質の翻訳は mRNA 上の開始コドンから始まって終止コドンで終了する。このアミノ酸配列に翻訳されうる連続した塩基配列をオープンリーディングフレーム(ORF)と呼ぶ。
- 3 mRNA の 5'末端と 3'末端の翻訳されない領域を非翻訳領域(Untranslated region : UTR)と呼ぶ。
- 4 リボソーム上でのタンパク質の翻訳はアミノ酸のカルボキシル(C)末端からアミノ(N)末端の方向に向かって進行する。

問 8

細胞小器官（細胞内小器官）に関する次の記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 核は 2 枚の脂質二重膜によって細胞質と分けられており、遺伝情報が格納されている。核内と細胞質の物質のやりとりは核膜孔を通じて行われる。
- 2 ミトコンドリアはエネルギーの産生を担っている。ミトコンドリアは独自のゲノムを持つが、ミトコンドリアを構成するタンパク質の一部は核にコードされている。
- 3 液胞は代謝物や不要物質の貯蔵、浸透圧の調整、不要物質の細胞内消化の機能を担う。これは動物細胞で発達しており、細胞の大半を占める場合がある。
- 4 リソソームは小胞構造で、内部に多くの消化酵素を含んでいる。リソソームの内部ではタンパク質や核酸、脂質の分解が行われており細胞内の物質リサイクルを担っている。

問 9

真核生物の特徴に関する以下の記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 核を有し、遺伝情報はゲノム DNA として核内に格納されている。
- 2 ヌクレオソームは、真核生物に見られるヒストンに DNA が巻き付いた構造である。
- 3 核内で転写された mRNA は細胞質に輸送された後に翻訳される。
- 4 RNA ポリメラーゼは I、II、III の 3 種類存在し、mRNA は RNA ポリメラーゼ III によって転写される。

問 10

原核生物に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 原核生物は核、ミトコンドリア、小胞体などの細胞小器官（細胞内小器官）を持たない。
- 2 原核生物は一般に単細胞生物である。
- 3 原核生物のゲノムは一般に直鎖状であり、有糸分裂を行う。
- 4 原核生物は真正細菌と古細菌に大別される。

問 11

RNA に関する以下の記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 リボソームはタンパク質と RNA の複合体である。このリボソームを構成する RNA をリボソーム RNA(rRNA)と呼ぶ。
- 2 RNA はしばしばその塩基が化学修飾を受けている。この中でユビキチン化は RNA 分解のシグナルとして機能する。
- 3 マイクロ RNA(miRNA)は 18 塩基から 25 塩基の短い RNA である。RISC 複合体は miRNA を取り込むとその塩基配列依存的に標的 mRNA に結合する。この結果、標的 mRNA は切断され、あるいは翻訳が抑制される。
- 4 長鎖ノンコーディング RNA(long noncoding RNA あるいは lncRNA)は、mRNA と同じ機構で合成されるがタンパク質をコードしない比較的長い (200 塩基～) RNA である。lncRNA の例として X 染色体の不活化に関わる *Xist* などが知られている。

問 12

遺伝子組換えに関して述べた以下の記述のうちでもっとも不適切なものを選択肢の中から一つ選べ。

- 1 遺伝子組換え生物はもともと天然に存在しないため、人類の健康や自然環境への影響が未知数である。したがって遺伝子組換え生物を生み出す行為とその産物は適切なルールの下に管理される必要がある。
- 2 siRNA とよばれる人工合成した低分子の二本鎖 RNA を細胞内に導入すると、この RNA と相同な DNA を切断する活性を持った RITS 複合体が形成されて標的遺伝子が破壊される。そのため siRNA は遺伝子のノックアウトに有効である。
- 3 化学合成した複数の相補的なオリゴヌクレオチドを貼り合わせ、ギャップやニックを DNA ポリメラーゼや DNA リガーゼで埋めることによって、より長い任意の DNA 配列を人工合成することが可能となる。
- 4 CRISPR-Cas9 システムを用いると人工的にデザイン・合成したガイド RNA の配列特異性を利用して任意のゲノム領域に DNA の二本鎖切断を導入することが可能となる。この結果、遺伝子の破壊や編集を容易に実現することが可能になる。

(問 13・問 14)

以下は真核生物のとあるゲノム領域で、下線部が mRNA に転写される。

【ゲノム領域】

5'-ATATATTCTCAATAAGGTATGCTTGTAGCATGGTAAGTGTTTTAAACGAGAAAT-3'

これをふまえて、以下の問 13 および問 14 に答えよ。

問 13

合成される mRNA の構造としてもっとも適切なものを選択肢の中から一つ選べ。ただし、m7G、m7A はそれぞれ 7-methylguanosine、7-methyladenosine を意味し、3 リン酸を介して mRNA の 5'末端に連結されていることを想定している。

- 1 5'-m7G-AAUAAGGUAUGCUUGUAGCAUGGUAAGUGUUUUAAAAAAAAAAAAAAAAAAAA-3'
- 2 5'-m7A-AAUAAGGUAUGCUUGUAGCAUGGUAAGUGUUUUUUUUUUUUUUUUUUUUUUUUUU-3'
- 3 5'-m7G-AAUAAGGUAUGCUUGUAGCAUGGUAAGUGUUUUUUUUUUUUUUUUUUUUUUUUUU-3'
- 4 5'-m7A-AAUAAGGUAUGCUUGUAGCAUGGUAAGUGUUUUAAAAAAAAAAAAAAAAAAAA-3'

問 14

問 13において転写された mRNA 上のオープンリーディングフレームは翻訳される。このタンパク質の性質についてもっとも適切なものを選択肢の中から一つ選べ。回答にあたっては次ページの遺伝暗号表を参照せよ。

- 1 非極性のアミノ酸のみである。
- 2 生理条件下では一部のアミノ酸の側鎖が正に荷電している。
- 3 生理条件下では一部のアミノ酸の側鎖が負に荷電している。
- 4 分子内でジスルフィド結合を形成する可能性がある。

【遺伝暗号表】

| 1 文字目 | 3 文字目 | 2 文字目 |     |      |     |     |     |     |
|-------|-------|-------|-----|------|-----|-----|-----|-----|
|       |       | U     | C   | A    | G   |     |     |     |
| U     | U     | Phe   | Ser | Tyr  | Cys |     |     |     |
|       | C     |       |     | Stop | Trp |     |     |     |
|       | A     | Leu   |     |      |     | Pro | His | Arg |
|       | G     |       |     |      |     |     |     |     |
| U     | Gln   |       |     |      |     |     |     |     |
| C     |       |       |     |      |     |     |     |     |
| A     |       |       |     |      |     |     |     |     |
| G     |       |       |     |      |     |     |     |     |
| A     | U     | Ile   | Thr | Asn  | Ser |     |     |     |
|       | C     |       |     | Lys  | Arg |     |     |     |
|       | A     | Met   |     |      |     |     |     |     |
|       | G     |       |     |      |     |     |     |     |
| G     | U     |       | Val | Ala  | Asp | Gly |     |     |
|       | C     |       |     |      | Glu |     |     |     |
|       | A     |       |     |      |     |     |     |     |
|       | G     |       |     |      |     |     |     |     |

問 15

エピゲノムに関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ゲノム配列が同一でも DNA のメチル化修飾部位が異なる場合、エピゲノムは異なるとみなされる。
- 2 同一個体内でも臓器によってエピゲノムは異なる。
- 3 食事やストレスなどの外的要因でエピゲノムは変化する。
- 4 エピゲノム解析の手法のひとつとしてエクソーム解析がある。

問 16

DNA の複製に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 複製では親鎖の 2 重らせんが一本鎖に分かれ、それぞれの鎖を鋳型として相補的な塩基配列を持つ DNA 鎖が新たに合成される。これを半保存的複製という。
- 2 DNA ポリメラーゼは、鋳型となる DNA 鎖上を 3'末端→ 5'末端の方向に移動しながら、鋳型 DNA の塩基配列に相補的なデオキシリボヌクレオチドを 5'末端→ 3'末端方向に結合していく。
- 3 リーディング鎖では、複製フォークで鋳型 DNA が一本鎖に開かれるにつれて少しずつ新しい鎖が合成されるため、短い DNA 鎖（岡崎フラグメント）となる。
- 4 原核生物は複製開始点が染色体あたり 1 カ所のみ存在するのに対して、真核生物のゲノム DNA の複製開始点は複数存在する。

問 17

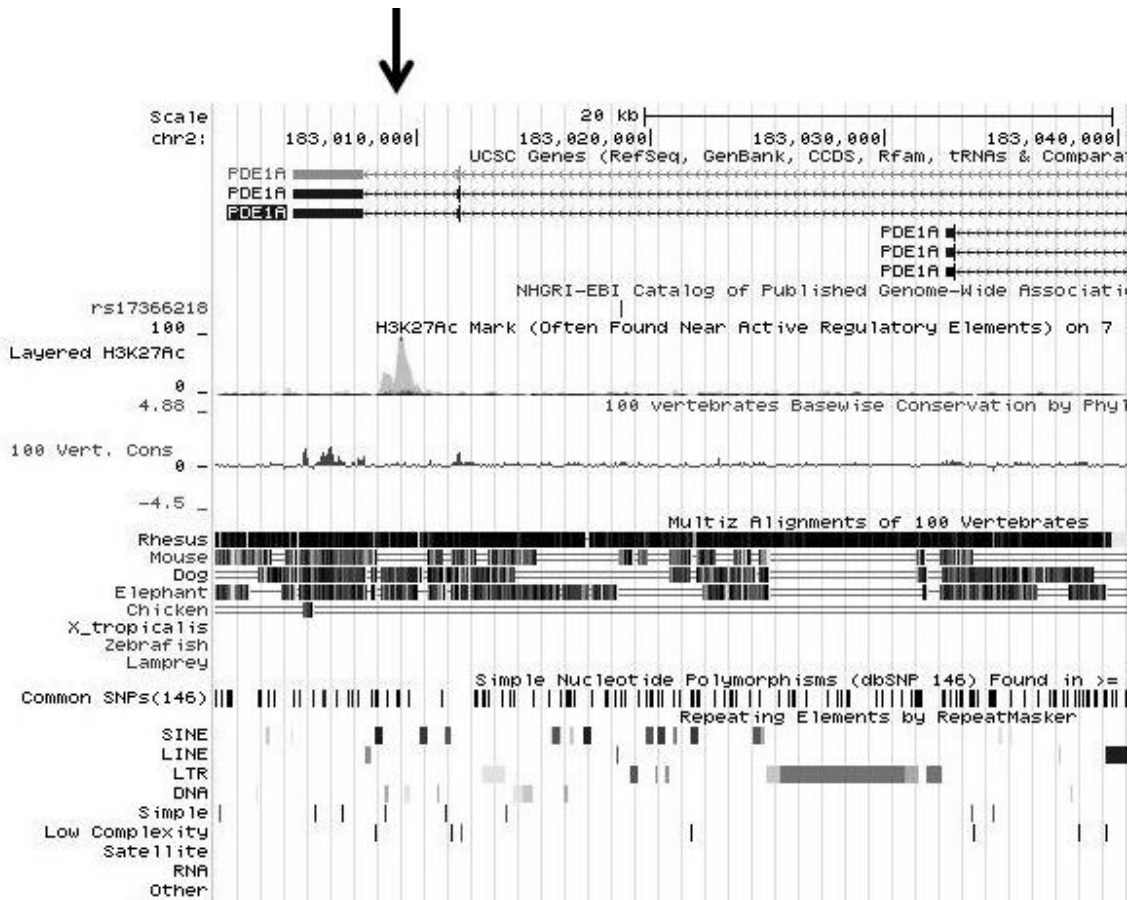
逆転写に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 セントラルドグマが発表された後に逆転写のシステムが新たに発見された。
- 2 RNA ウイルスの中で逆転写酵素を持たないものはレトロウイルスと呼ばれる。
- 3 cDNA の合成は、RNA を鋳型として逆転写酵素で相補的な DNA を合成することで行われる。
- 4 ヒトゲノムに含まれる LTR 型レトロトランスポゾンとは両側に LTR(long terminal repeat)と呼ばれる DNA 配列をもち、逆転写によりゲノム中を転移する。



問 18

以下に示すデータは、UCSC genome browser で可視化したヒトゲノム上のある領域である。図の矢印の領域について、図から読み取れる情報としてもっとも不適切なものを選択肢の中から一つ選べ。



- 1 PDE1A gene の転写開始点付近である。
- 2 PDE1A gene は複数のアイソフォームを持つ。
- 3 H3K27ac のヒストン修飾が入っていることから、遺伝子発現制御機能に関連する領域であると考えられる。
- 4 サル、イヌでは保存されているが、マウスでは保存されていない領域である。

問 19

DNA マイクロアレイに関する以下の文章の( a )～( d )に入る語句の組み合わせとしてもっとも適切なものを選択肢の中から一つ選べ。

DNA マイクロアレイは基板に固定された多数の( a )を用いて、サンプル中の DNA、もしくは RNA の量を測定する技術である。遺伝子発現量の測定のほか、( b )やジェノタイピングにも利用される。ジェノタイピング解析には SNP アレイや( c )が利用され、未知の遺伝子の検出には全ゲノム領域をカバーする( d )が用いられる。

- |   |          |                  |             |              |
|---|----------|------------------|-------------|--------------|
| 1 | (a) リガンド | (b) ChIP-on-chip | (c) メチル化アレイ | (d) エキソンアレイ  |
| 2 | (a) リガンド | (b) ChIP-seq     | (c) メチル化アレイ | (d) タイリングアレイ |
| 3 | (a) プローブ | (b) ChIP-on-chip | (c) CNV アレイ | (d) タイリングアレイ |
| 4 | (a) プローブ | (b) ChIP-seq     | (c) CNV アレイ | (d) エキソンアレイ  |

問 20

PCR(polymerase chain reaction)に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 PCR 法では、反応液の温度を変化させ、二本鎖 DNA の変性、プライマーのアニーリング、相補鎖 DNA の合成、二本鎖の解離を繰り返すことによって DNA 断片が増幅される。
- 2 PCR 法の反応には好熱菌由来の耐熱性の DNA ポリメラーゼがよく使われる。
- 3 RT-PCR(Reverse Transcription-PCR)法は細胞や組織から RNA を抽出したあとに、プライマー DNA、逆転写酵素により cDNA を合成し、この cDNA をテンプレートにして PCR を行う方法である。
- 4 DNA の定量を目的とした定量 PCR 法では、PCR 効率をできるだけ高めるため、増幅されやすい反復配列を含むプライマーを設計する。

問 21

$1.0 \times 10^{10}$  以上である浮動小数点数  $x$  に対し、桁落ちの問題がもっとも大きくなるとされる計算式を選択肢の中から一つ選べ。

- 1  $\frac{1}{-x - \sqrt{x^2 + 1}}$
- 2  $x - \sqrt{x^2 + 1}$
- 3  $\frac{-1}{x + \sqrt{x^2 + 1}}$
- 4  $x + \sqrt{x^2 - 1}$

問 22

プログラミング言語の関数呼び出しに関する以下の文章およびコードの中で、a から c に入る語句の組み合わせとしてもっとも適切なものを選択肢の中から一つ選べ。

C 言語の関数呼び出しにおける引数の受け渡しは、a 渡しで行われる。関数の呼び出し元の変数の値を関数の中で変更するには、ポインタを利用し、変数のアドレスを関数に渡せばよい。例えば、以下のプログラムでは main 関数内の変数  $x$  は increment 関数内で変更され、標準出力に 1 が出力される。

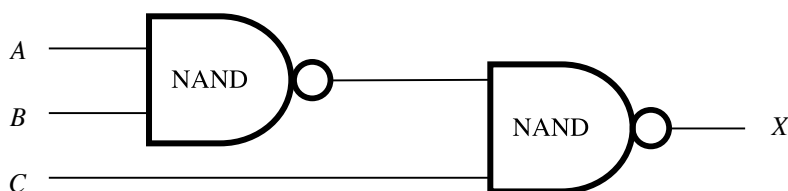
```
#include <stdio.h>
void increment(int b) {
    (*i)++;
}
int main(void) {
    int x = 0;
    increment(c);
    printf("%d\n", x);
    return 0;
}
```

- |   |        |        |        |
|---|--------|--------|--------|
| 1 | (a) 値  | (b) &i | (c) &x |
| 2 | (a) 値  | (b) *i | (c) *x |
| 3 | (a) 値  | (b) *i | (c) &x |
| 4 | (a) 参照 | (b) &i | (c) x  |

問 23

下記の図は、2つの論理素子を接続した論理回路を表現している。この回路では  $A$ ,  $B$ ,  $C$  が入力であり、 $X$  が出力であるとする。 $X$  に対応する論理式として間違っているものを選択肢の中から一つ選べ。

なお、「 $PQ$ 」は $P$ と $Q$ の論理積（「 $P$  and  $Q$ 」あるいは「 $P \wedge Q$ 」とも書く）、「 $P + Q$ 」は $P$ と $Q$ の論理和（「 $P$  or  $Q$ 」あるいは「 $P \vee Q$ 」とも書く）、「 $\bar{P}$ 」は $P$ の否定（「not  $P$ 」あるいは「 $\neg P$ 」とも書く）を表す。



- 1  $X = \overline{\overline{(AB)C}}$
- 2  $X = (AB) + (\bar{C}A) + \overline{(C + A)}$
- 3  $X = (AB) + \bar{C}$
- 4  $X = \overline{ABC}$

問 24

以下の中から、IEEE754 の 2 進形式単精度浮動小数点数によって誤差なく表現することができない 10 進数小数を選択肢の中から一つ選べ。

- 1 0.1
- 2 0.75
- 3 1.5
- 4 10.0

問 25

計算量の記述によく用いられるビッグオー記法に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。ただしここでは関数はすべて自然数を非負実数へ写すとする。

- 1 関数  $f$  が  $f(n) = O(n)$  を満たしても  $f(n) > n^2$  となる  $n$  が存在する場合がある。
- 2 関数  $f$  が  $f(n) = O(n)$  を満たすならば十分大きい  $n$  に対しては  $f(n) < n^2$  である。
- 3 関数  $f$  が  $f(n) = O(n^2)$  を満たすならば十分大きい  $n$  に対しては  $n < f(n)$  である。
- 4 任意の関数  $f$  と  $g$  に対して  $f(n) + g(n) = O(\max\{f(n), g(n)\})$  である。

問 26

一列に並んだデータを格納するデータ構造として連結リストを用いた場合と配列を用いた場合とを考える。次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。ただし、以下では  $n$  はデータの要素数を表すとする。

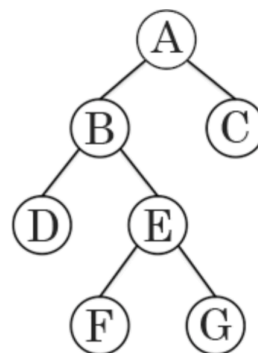
- 1 整数値  $k$  ( $1 \leq k \leq n$ ) を受け取り  $k$  番目の要素を削除する操作は、連結リストでは  $O(1)$  時間で可能であるが、配列では順序を保存する必要がなかったとしても  $O(1)$  時間では不可能である。
- 2 整数値  $k$  ( $1 \leq k \leq n$ ) を受け取り  $k$  番目の要素を求める操作は、連結リストでは  $O(n)$  時間、配列では  $O(1)$  時間で可能である。
- 3 先頭から順番にすべての要素を参照する場合、一般に配列の方が高速である。
- 4 格納する対象のデータが同じでも、連結リスト、配列どちらを用いるかによって使用する総メモリ量が異なる。

問 27

以下の擬似コードは、キュー(擬似コード内の  $Q$ )を用いて木を巡回し、ノード  $v$  以下の子孫ノードを  $v$  を含めすべて `output` 関数で出力するプログラムを表す。

`traverse(v) :`

```
{
  enqueue v to Q
  while (Q is not empty) {
    v ← dequeue a node from Q
    output(v)
    for w ∈ children(v) {
      enqueue w to Q
    }
  }
}
```

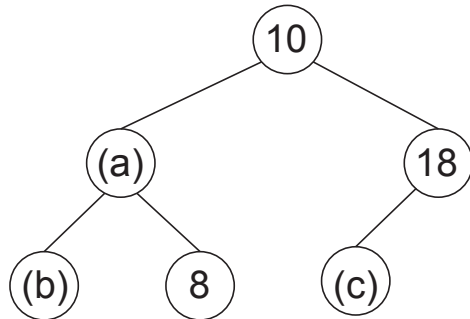


ただし `enqueue` と `dequeue` はキュー  $Q$  に対する `enqueue` 操作と `dequeue` 操作を表す。また、「for  $w \in \text{children}(v) \dots$ 」は、ノード  $v$  の子のノードそれぞれに対し左側の子のノードから順に計算を行うことを表すものとする。このとき、図の木に対して `traverse(A)` を実行した際の出力の組み合わせとして正しいものを選択肢の中から一つ選べ。

- 1 A, B, C, D, E, F, G
- 2 A, C, B, E, G, F, D
- 3 A, B, D, E, F, G, C
- 4 D, F, G, E, B, C, A

問 28

数列(10, 6, 18, 3, 8, 15)を入力として二分探索木を作成した所、以下の図のようになった。この図の ( a )から( c )に入る数字の組み合わせとして正しいものを選択肢の中から一つ選べ。



- |   |       |        |        |
|---|-------|--------|--------|
| 1 | (a) 3 | (b) 6  | (c) 15 |
| 2 | (a) 6 | (b) 15 | (c) 6  |
| 3 | (a) 6 | (b) 3  | (c) 15 |
| 4 | (a) 3 | (b) 15 | (c) 6  |

問 29

問 28 の図に関して、以下の問いに答えよ。

二分木においてその中間順は、二分木のルートノード  $t$  に対して、以下の擬似コードを用いて `traverse(t)` を計算した際に、`output()`関数によって出力されるノードの順序である。ただし、`left_child(v)` は  $v$  の子ノードのうち左側のノード（なければ `nil`）、`right_child(v)`は  $v$  の子ノードのうち右側のノード（なければ `nil`）を返す関数である。

`traverse (v) :`

```

{
  if (v ≠ nil) {
    traverse(left_child(v))
    output(v)
    traverse(right_child(v))
  }
}
  
```

このとき、問 28 中の図の木に対し、その中間順として出力される数字の順序として正しいものを選択肢の中から一つ選べ。

- |   |                     |
|---|---------------------|
| 1 | 3, 6, 8, 10, 15, 18 |
| 2 | 3, 8, 6, 15, 18, 10 |
| 3 | 10, 6, 3, 8, 18, 15 |
| 4 | 10, 6, 18, 3, 8, 15 |

問 30 (※訂正があります。HP に掲載しておりますのでご参照下さい。以下は当日出題した問題です)

2 本の DNA 配列のアラインメントは動的計画法を用いることにより計算することができる。

一致スコアを 1、不一致スコアを-2、ギャップスコアを-1 とした時、DNA 配列 ATGC と ACTC のアラインメントを計算するため、下のテーブルの空欄( ア )( イ )を埋める組み合わせとして正しいものを選択肢の中から一つ選べ。

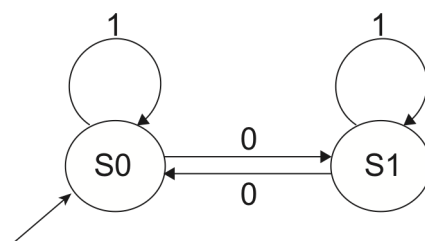
|   | A  | T   | G  | C   |
|---|----|-----|----|-----|
| A | 0  | -1  | -2 | -3  |
| C | -1 | (ア) | 0  | -1  |
| T | -2 | 0   | -1 | -2  |
| C | -3 | -1  | 2  | (イ) |
|   | -4 | -2  | 1  | 0   |

- 1 (ア) 1 (イ) -3
- 2 (ア) -2 (イ) 1
- 3 (ア) -2 (イ) 0
- 4 (ア) 1 (イ) 1

問 31

下図の決定性有限オートマトンは2つの状態S0、S1を持つ。S0は初期状態である。入力数列は0または1からなる。このオートマトンに関する記述で間違っているものを選択肢の中から一つ選べ。

| 状態遷移表 |   | 状態 |    |
|-------|---|----|----|
|       |   | S0 | S1 |
| 入力    | 0 | S1 | S0 |
|       | 1 | S0 | S1 |



- 1 入力数列が 01001 のとき、状態は S1 となる。
- 2 入力数列中に 0 が奇数個ある時、状態は S1 となる。
- 3 入力数列中に偶数個の 1 と奇数個の 0 が含まれるとき、状態は S0 となる。
- 4 入力数列が 10100110 のとき、状態は S0 となる。

問 32

データ格納方式としてスタックを用いる。アルファベットの文字データをスタックに出し入れする。データの操作を

Push A  
Push B  
Push C  
Push D  
Pop  
Pop  
Push B  
Push D  
Pop  
Pop  
Pop

のように行った時、最後の Pop 操作で取り出されるデータとして正しいものを選択肢の中から一つ選べ。ただし、Push はデータを追加する操作、Pop はデータを取り出す操作である。

- 1 A
- 2 B
- 3 C
- 4 D

問 33

すでにソートされている長さ  $n$  の相異なる数字からなる数列に対してクイックソートをもう一度かけることを考える。このクイックソートにおいてピボット(pivot)を必ず先頭の数字を選ぶものとした場合の計算量についての記述のうち、もっとも適切なものを選択肢の中から一つ選べ。

- 1 常に $O(n)$ 時間で終了する。
- 2 常に $O(n \log n)$ 時間で終了する。どのような入力（すでにソートされている長さ  $n$  の相異なる数字からなるどのような数列）に対しても $O(n)$ 時間で終了することはない。
- 3 常に $O(n^2)$ 時間で終了する。どのような入力（すでにソートされている長さ  $n$  の相異なる数字からなるどのような数列）に対しても $O(n \log n)$ 時間で終了することはない。
- 4 アルゴリズムが停止するとは限らない。



問 34

下記の表に対して、次に示す SQL 文を実行する。結果として得られるデータの行数として正しいものを選択肢の中から一つ選べ。

SELECT \* FROM 測定結果,サンプル量  
WHERE 測定結果.サンプル ID=サンプル量.サンプル ID

測定結果

| 患者 ID | サンプル ID | 測定値 A | 測定値 B |
|-------|---------|-------|-------|
| AR397 | 10205   | 41.7  | 3.5   |
| AR397 | 10206   | 41.9  | 6.9   |
| ZN422 | 273     | 37.5  | 2.1   |
| JH032 | 4423    | 52.8  | 144.5 |
| RR941 | 551     | 30.9  | 0.0   |

サンプル量

| 患者 ID | サンプル ID | 量    |
|-------|---------|------|
| AR397 | 10206   | 30.0 |
| AR397 | 10207   | 30.0 |
| AR397 | 10209   | 30.0 |
| RR941 | 560     | 10.0 |
| RR941 | 569     | 30.0 |

- 1 1 件
- 2 2 件
- 3 3 件
- 4 4 件

問 35

標準正規分布  $N(0,1)$  の密度関数  $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$  の累積密度関数  $F(x) = \int_{-\infty}^x f(t)dt$  について、正しいものを選択肢の中から一つ選べ。

- 1  $F(0) = 0$
- 2  $F(0) = 0.5$
- 3  $F(0) = 1$
- 4  $F(0) = 1/(2\pi)$

問 36

以下の ( a ) ~ ( c ) に入れる語としてもっとも適切なものを選択肢の中から一つ選べ。

期待値  $\lambda$  のポアソン分布  $p$  の分散は  $\lambda$  になることが知られている。この分布  $p$  から統計的独立に生成された  $n$  個の標本値の標本平均の分布は平均( a )、分散( b )の( c )に漸近する。

- |   |                   |                   |           |
|---|-------------------|-------------------|-----------|
| 1 | (a) $\lambda / n$ | (b) $\lambda$     | (c) ガンマ分布 |
| 2 | (a) $\lambda$     | (b) $\lambda / n$ | (c) ガンマ分布 |
| 3 | (a) $\lambda / n$ | (b) $\lambda$     | (c) 正規分布  |
| 4 | (a) $\lambda$     | (b) $\lambda / n$ | (c) 正規分布  |

問 37

3つの事象 A、B、C に対して、B かつ C の条件下で A が生起する確率  $P(A|B,C)$  の満たす等式として正しいものを選択肢の中から一つ選べ。

- |   |                                |
|---|--------------------------------|
| 1 | $P(A B,C) = P(A,B C) / P(B C)$ |
| 2 | $P(A B,C) = P(A,B C) / P(C B)$ |
| 3 | $P(A B,C) = P(B,C) / P(B A,C)$ |
| 4 | $P(A B,C) = P(B,C) / P(B,C A)$ |

問 38

立方体のさいころを投げる試行を 2 回行う。どの目も  $1/6$  の確率で出るとし、2 回の試行は互いに独立であるとする。この試行に関する記述として、間違っているものを選択肢の中から一つ選べ。

- |   |                                         |
|---|-----------------------------------------|
| 1 | 2 回目の数字が 1 回目の数字よりも大きくなる確率は $1/2$ である。  |
| 2 | 1 回目と 2 回目のサイコロの目の値が等しくなる確率は $1/6$ である。 |
| 3 | 2 回のサイコロの目の値の和の期待値は 7 である。              |
| 4 | 2 つの目の値の差が 1 となる確率はその差が 2 となる確率よりも大きい。  |

問 39

以下の( a )、( b )に入れる語として正しいものを選択肢の中から一つ選べ。

検索システムの性能を測る指標として、精度(Precision)、再現率(Recall)、F 値(F-measure あるいは F-score)がよく用いられる。F 値は精度と再現率の調和平均で定義されている。すなわち、精度を  $P$ 、再現率を  $R$  で表すとする、F 値は( a )で与えられる。システムの性能がよいほど、F 値は( b )に近づく。

- |   |                   |       |
|---|-------------------|-------|
| 1 | (a) $(P+R) / 2$   | (b) 1 |
| 2 | (a) $(P+R) / 2$   | (b) 0 |
| 3 | (a) $2PR / (P+R)$ | (b) 1 |
| 4 | (a) $2PR / (P+R)$ | (b) 0 |

問 40

2 次元ユークリッド空間  $E$  上のユークリッド距離  $D(x, y)$  について、間違っているものを選択肢の中から一つ選べ。

- 1  $D(x, y) < 0$  となる  $x, y$  が  $E$  上に存在する。
- 2  $E$  上の 2 点  $x, y$  に対して、 $D(x, y) = 0$  ならば、 $x=y$  が成り立つ。
- 3  $E$  上の任意の 3 点  $x, y, z$  に対して、 $D(x, y) \leq D(x, z) + D(z, y)$  が成り立つ。
- 4  $E$  上の任意の 2 点  $x, y$  に対して、 $D(x, y) = D(y, x)$  が成り立つ。

問 41

国際的な公共の分子生物学データベースとデータに関する組合せでもっとも不適切なものを選択肢の中から一つ選べ。

| 選択肢 | データベース       | データ    |
|-----|--------------|--------|
| 1   | Pfam         | 配列モチーフ |
| 2   | PubMed       | 文献情報   |
| 3   | ArrayExpress | 遺伝子発現  |
| 4   | SCOP         | 遺伝子多型  |

問 42

2 本のアミノ酸配列でグローバルアラインメントを求めたところ、図のような結果を得た。このアラインメントのスコアを、図中のアミノ酸置換行列（一部）を用いて求め、選択肢の中からもっとも適した値を一つ選べ。なお、ギャップペナルティは 1（スコアから減算する）とする。

GKRFD-C  
 | | : |  
 G-RFEVC

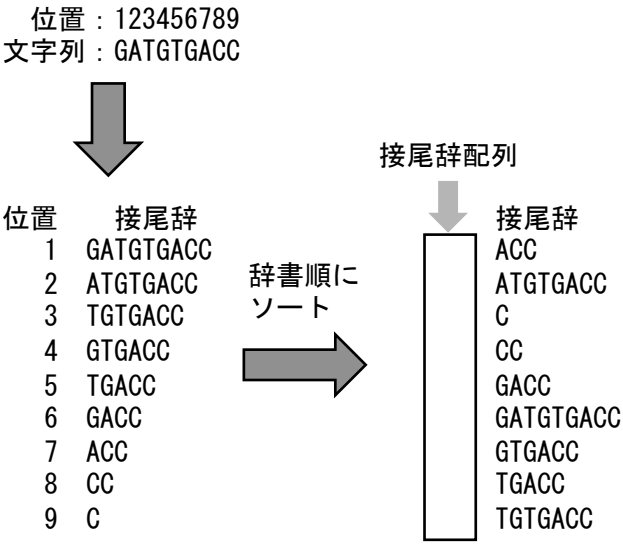
アミノ酸置換行列の一部

|   | R  | D  | C  | E  | G  | K  | F  | V  |
|---|----|----|----|----|----|----|----|----|
| R | 6  | -1 | -4 | -1 | -3 | 3  | -4 | -4 |
| D | -1 | 4  | -5 | 3  | 1  | 0  | -6 | -2 |
| C | -4 | -5 | 12 | -5 | -3 | -5 | -4 | -2 |
| E | -1 | 3  | -5 | 4  | 0  | 0  | -5 | -2 |
| G | -3 | 1  | -3 | 0  | 5  | -2 | -5 | -1 |
| K | 3  | 0  | -5 | 0  | -2 | 5  | -5 | -2 |
| F | -4 | -6 | -4 | -5 | -5 | -5 | 9  | -1 |
| V | -4 | -2 | -2 | -2 | -1 | -2 | -1 | 4  |

- 1 28
- 2 31
- 3 33
- 4 35

問 43

下図では、配列 GATGTGACC を接尾辞配列へ変換する過程を示している。空欄の接尾辞配列としてもっとも適したものを選択肢の中から一つ選べ。



- 1 1,2,3,4,5,6,7,8,9
- 2 7,2,9,8,6,1,4,5,3
- 3 7,2,8,9,6,1,4,5,3
- 4 2,7,9,8,4,1,6,3,5

問 44

配列解析の方法と、それと関連の深いアルゴリズム・データ構造の組み合わせとして、もっとも不適切なものを選択肢の中から一つ選べ。

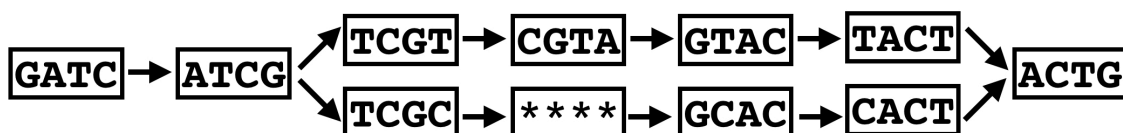
| 選択肢 | 配列解析の方法                           | アルゴリズム・データ構造                  |
|-----|-----------------------------------|-------------------------------|
| 1   | ホモロジー検索                           | ハッシュテーブル                      |
| 2   | マルチプルアラインメント                      | 逐次改善法                         |
| 3   | ショートリード（100-150bp）の参照ゲノム配列へのマッピング | バローズ・ホイーラー（Burrows-Wheeler）変換 |
| 4   | 配列モチーフ検索                          | ホモロジーモデリング                    |

問 45

以下の文章を埋める最適な語句、配列となる組み合わせを選択肢の中から一つ選べ。

通常、配列解析で扱われるデータにはシーケンシングエラーや反復配列が含まれている。そのため、 $k$ -merを頂点とするド・ブラン(de Bruijn)グラフを用いたショートリードのアセンブルでは分岐を含んだグラフが生じる。図はパラメータ  $k$  が 4 の場合の、あるド・ブラングラフを表す。図中の「\*\*\*\*」の配列は( a )となる。また、生じる配列は「GATCGTACTG」と「( b )」となる。このような場合、( c )の出現頻度やペアエンドの情報を用いてより確からしいパスを決定するなどして最終的なアセンブリを構築する。

- |   |          |                |              |
|---|----------|----------------|--------------|
| 1 | (a) CGTA | (b) GATCGTAGTG | (c) $k$ -mer |
| 2 | (a) CGTA | (b) GATCGCACTG | (c) コドン      |
| 3 | (a) CGCA | (b) GATCGCACTG | (c) $k$ -mer |
| 4 | (a) CGCA | (b) GATCGCAGTG | (c) コドン      |



問 46

ある塩基配列に対して BLAST を用いて相同性検索を行った結果は Score = 200、Expect =  $4e-20$  であった。この結果の解釈を述べた以下の文中において、空欄( a )、( b )にあてはまる語句の組合せとして、もっとも適切なものを選択肢の中から一つ選べ。

データベース中に相同な配列が存在しない場合でも、検索結果に( a )のスコアが少なくとも 1 つ含まれる確率は、およそ( b )である。

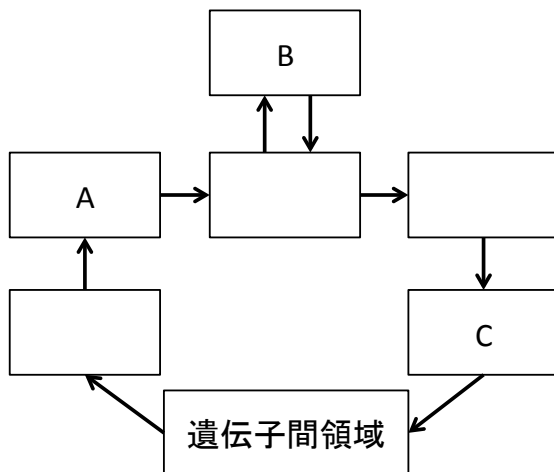
- |   |            |                             |
|---|------------|-----------------------------|
| 1 | (a) 200 以上 | (b) $4 \times 10^{-20}$     |
| 2 | (a) 200 以下 | (b) $4 \times 10^{-20}$     |
| 3 | (a) 200 以上 | (b) $1 - 4 \times 10^{-20}$ |
| 4 | (a) 200 以下 | (b) $1 - 4 \times 10^{-20}$ |

問 47

下図は真核生物の遺伝子構造のモデルを表している。ただし、ここでは簡単のため遺伝子は一方の鎖のみにコードされており、図の矢印は遺伝情報が読み取られる方向に一致しているものとする。また、図の各枠内には以下にリストされた語句のいずれかが入ることとする。図の A、B、C に入る組合せとして、もっとも適切なものを選択肢の中から一つ選べ。

【語句リスト】

イントロン、コーディングエクソン、3' UTR、5' UTR、プロモーター、ポリ A シグナル



| 選択肢 | A         | B          | C         |
|-----|-----------|------------|-----------|
| 1   | 5' UTR    | イントロン      | ポリ A シグナル |
| 2   | プロモーター    | イントロン      | 3' UTR    |
| 3   | 3' UTR    | コーディングエクソン | プロモーター    |
| 4   | ポリ A シグナル | コーディングエクソン | 5' UTR    |

問 48

次世代シーケンサから得られたショートリードのマッピングにおいては、クオリティ値が低い文字はシーケンスエラーの可能性があるので、文字列の照合時にミスマッチを無視したり、低く重みづけしたりすることがある。「クオリティ値 20 以下のときミスマッチを無視する」というルールの下で、以下のリード配列とリファレンス配列とを照合したときのミスマッチの個数はいくつになるか。選択肢の中から一つ選べ。なお、クオリティ値は、クオリティ値 + 33 のアスキー文字によって表現されている。

リファレンス配列    GCTTCTCCTAGC  
 リード配列            GAGTCTCCTAAC  
 クオリティ値        ##8AEFHIIJ/+

アスキー文字コード表（一部）

|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 数値 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 |
| 文字 | !  | "  | #  | \$ | %  | &  | '  | (  | )  | *  | +  | ,  | -  | .  | /  |
| 数値 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 |
| 文字 | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | :  | ;  | <  | =  | >  |
| 数値 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 |
| 文字 | ?  | @  | A  | B  | C  | D  | E  | F  | G  | H  | I  | J  | K  | L  | M  |

- 1   0 個
- 2   1 個
- 3   2 個
- 4   3 個



(問 49・問 50・問 51)

同じ長さの DNA 配列 2 本(*Seq1*, *Seq2*)を並べて、そのままギャップなしでアラインメントしたとき、局所的に最大スコアをとる領域 (開始位置 *Begin*、終了位置 *End*) とそのときのスコア(*MaxScore*)は、右図の擬似コードで示す関数 *UngappedLocalAlign* によって計算できる。ただし、DNA 配列 (文字列) *Seq* はインデックスが 1 から始まる配列 (アレイ) として格納されているとし、*Length(Seq)*はその長さを返す関数である。また、スコアは一致(*match*)が 1 点、不一致(*mismatch*)が-1 点とする。以下の問 49～問 51 に答えよ。

問 49

これに関して述べた以下の文章の空欄 X、Y、Z にあてはまる語句・値の組合せとしてもっとも適切なものを選択肢の中から一つ選べ。

このアルゴリズムは、配列アラインメントにおける ( X )法を簡略化したものと解釈できる。これを下図のアラインメントに適用したとき、7 文字目を処理した後の a の時点での *Score* の値は( Y )、また *Score* の値が最高となる点は( Z )となる。

|              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|--------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| <i>Seq1</i>  | C | G | A | G | T | G | A | G | G | T | A | T | G | A | C | G | G |
|              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| <i>Seq2</i>  | C | T | A | C | C | G | A | C | G | T | A | A | G | A | T | A | G |
| <i>Score</i> | 1 |   |   |   |   |   | a |   |   | b |   |   |   | c |   |   | d |

```

function UngappedLocalAlign(Seq1, Seq2)
  match = 1; mismatch = -1
  Score = MaxScore = 0
  Begin = End = p = 0
  for i = 1 to Length(Seq1) do
    if Seq1[i] == Seq2[i] then
      Score += match
    else
      Score += mismatch
    endif
    if Score ≤ 0 then
      Score = 0
      p = i
    endif
    if Score > MaxScore then
      MaxScore = Score
      Begin と End の値を設定 (※)
    endif
  end
  return (MaxScore, Begin, End)
  
```

| 選択肢 | X           | Y | Z |
|-----|-------------|---|---|
| 1   | ニードルマン・ブンシュ | 1 | a |
| 2   | ニードルマン・ブンシュ | 2 | b |
| 3   | スミス・ウォーターマン | 2 | c |
| 4   | スミス・ウォーターマン | 1 | d |

問 50

擬似コード中の (※) の箇所では、局所的に最大スコアをとる領域の開始位置 *Begin* と終了位置 *End* ( $Begin \leq End$ ) の設定を行う。ここに入る文としてもっとも適切なものを選択肢の中から一つ選べ。

- 1  $Begin = p+1; End = i$
- 2  $Begin = p; End = i+1$
- 3  $Begin = i; End = p+1$
- 4  $Begin = i+1; End = p$

問 51

UngappedLocalAlign を用いて、あるゲノム上の長い領域の DNA 配列 *Seq* に対して以下の処理を行った。ただし、SubString(*Seq*, *Begin*, *End*)は、文字列 *Seq* の開始位置 *Begin* から終了位置 *End* までの部分文字列を返す関数とする (文字列のインデックスは 1 から始まる)。この処理によって見つけようとしている領域はどのようなものか。もっとも適切なものを選択肢の中から一つ選べ。

```
w=3
Seq1 = SubString( Seq, 1, Length(Seq) - w )
Seq2 = SubString( Seq, 1+w, Length(Seq) )
(MaxScore, Begin, End) = UngappedLocalAlign( Seq1, Seq2 )
```

- 1 コード領域
- 2 パリンドローム (回文) 構造
- 3 散在性反復配列
- 4 タンデム (縦列) 反復配列

問 52

下図のようなプロファイルHMMがある。状態間の遷移確率は図中に示されている。  
状態M1、M2、M3 での文字の出力確率は、それぞれ以下のようにになっている。

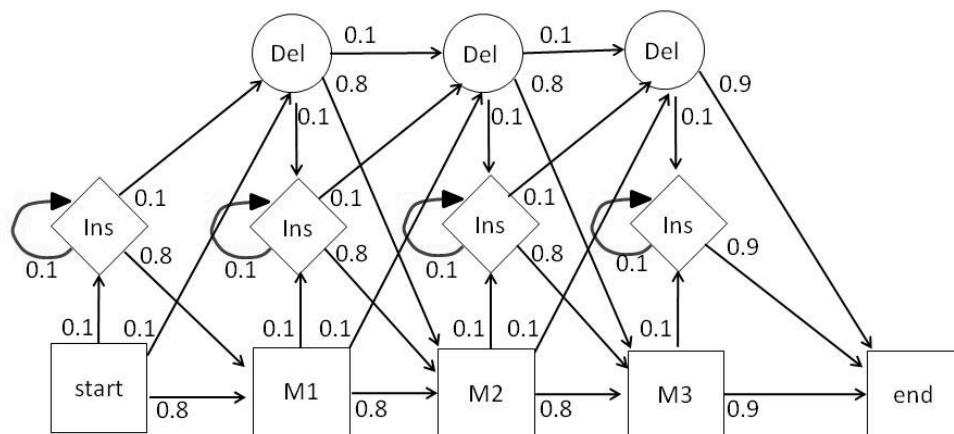
M1 :  $P(A)=0.0$   $P(T)=0.5$   $P(G)=0.5$   $P(C)=0.0$

M2 :  $P(A)=1.0$   $P(T)=0.0$   $P(G)=0.0$   $P(C)=0.0$

M3 :  $P(A)=0.0$   $P(T)=0.0$   $P(G)=0.9$   $P(C)=0.1$

またIns 状態では、位置によらず  $P(A)=P(T)=P(G)=P(C)=0.25$  である。

このHMM から、「start → T → A → C → G → end」という出力文字列が観測されたとする。以下の  
選択肢に示された状態遷移パスのうち、この出力文字列を発生させる確率をもっとも大きいものを一つ  
選べ。



- 1 start → Ins → M1 → M2 → M3 → end
- 2 start → M1 → Ins → M2 → M3 → end
- 3 start → M1 → M2 → Ins → M3 → end
- 4 start → M1 → M2 → M3 → Ins → end

問 53

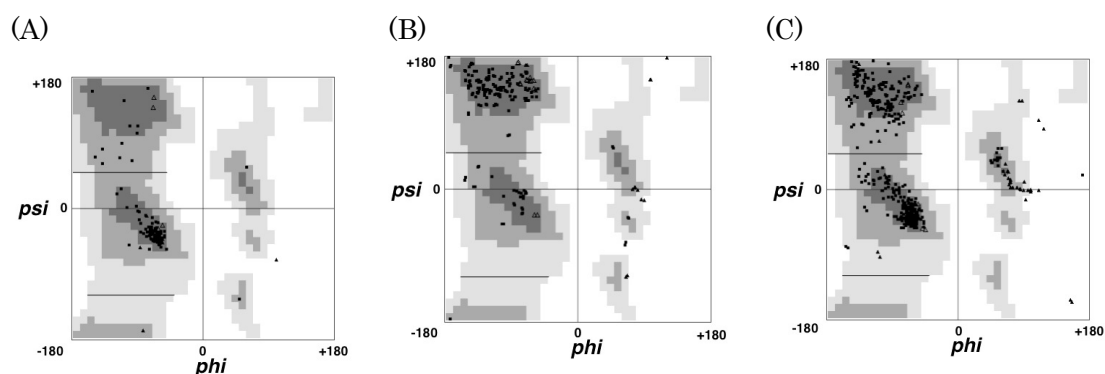
以下の天然変性タンパク質に関する記述の ( a ) ～ ( d ) に入る語句の組合せとして、もっとも適切なものを選択肢の中から一つ選べ。

タンパク質は特定の ( a ) を形成することにより、特異的な機能を発揮すると考えられてきた。しかしながら近年、特定の ( a ) を形成しない天然変性タンパク質の存在が明らかとなってきた。このようなタンパク質が含まれる割合は生物種により大きく変わり、特に ( b ) に数多く存在している。天然変性タンパク質は、( c ) 等に比較的多く観察される。またアミノ酸配列の特徴として、( d ) の含量が少なく、極性アミノ酸や荷電性アミノ酸が多く含まれている。これらの情報を利用して情報学的に天然変性領域を予測する手法が開発されている。

- |   |         |         |                     |            |
|---|---------|---------|---------------------|------------|
| 1 | (a)二次構造 | (b)真核生物 | (c)酸化還元酵素、リガーゼ      | (d)芳香族アミノ酸 |
| 2 | (a)三次構造 | (b)真核生物 | (c)転写因子、DNA 結合タンパク質 | (d)疎水性アミノ酸 |
| 3 | (a)二次構造 | (b)原核生物 | (c)転写因子、DNA 結合タンパク質 | (d)芳香族アミノ酸 |
| 4 | (a)三次構造 | (b)原核生物 | (c)酸化還元酵素、リガーゼ      | (d)疎水性アミノ酸 |

問 54

タンパク質のフォールドは、ポリペプチド鎖の結合の回転角で表現することができる。タンパク質主鎖の二面角(  $\Phi$ : *phi*,  $\Psi$ : *psi* )の二次元分布図はラムチャンドランマップ (ラムチャンドランプロット) と呼ばれ、フォールドを表現したり評価したりする手法として広く利用されている。以下の 3 つのラムチャンドランマップ A、B、C は、典型的なスーパーフォールドであるグロビンフォールド、免疫グロブリンフォールド、TIM バレルフォールドから作成したものである。それぞれのスーパーフォールドとラムチャンドランマップの組合せとして、もっとも適切なものを選択肢の中から一つ選べ。



- |   |                 |                 |                 |
|---|-----------------|-----------------|-----------------|
| 1 | (A)グロビンフォールド    | (B)TIM バレルフォールド | (C)免疫グロブリンフォールド |
| 2 | (A)免疫グロブリンフォールド | (B)グロビンフォールド    | (C)TIM バレルフォールド |
| 3 | (A)グロビンフォールド    | (B)免疫グロブリンフォールド | (C)TIM バレルフォールド |
| 4 | (A)TIM バレルフォールド | (B)免疫グロブリンフォールド | (C)グロビンフォールド    |

### 問 55

タンパク質の内部に埋もれたアミノ酸と露出したアミノ酸についての記述として、もっとも不適切なものを選択肢の中から一つ選べ。ただし、解答に際しては水溶性タンパク質と膜タンパク質の違いも考慮すること。

- 1 酵素の活性に直接関わる重要なアミノ酸であってもタンパク質内部に埋もれることがある。
- 2 タンパク質を構成するアミノ酸配列上で、タンパク質内部に位置するアミノ酸ほど保存度が高い傾向にある。
- 3 分子表面に位置していても、機能的制約がかかっているアミノ酸の保存度は高い傾向にある。
- 4 タンパク質の内側に埋もれたアミノ酸は疎水的、一方、露出したアミノ酸は親水的な傾向にある。

### 問 56

以下の図は32組の相同なタンパク質ペアについての配列一致度とRoot Mean Square Deviation

(RMSD: 根平均二乗偏差) の関係を示したものである。この図から読み取れることとして、もっとも不適切なものを選択肢の中から一つ選べ。

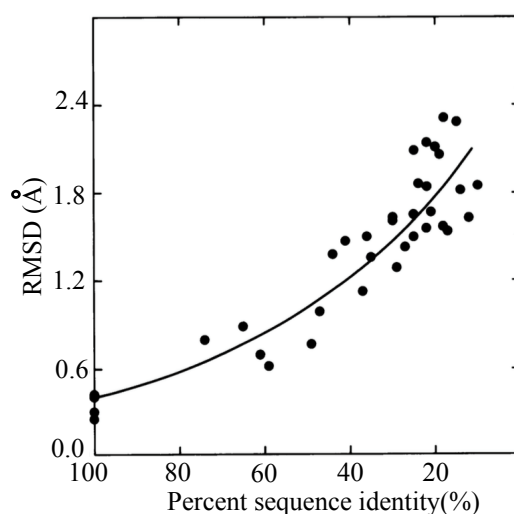


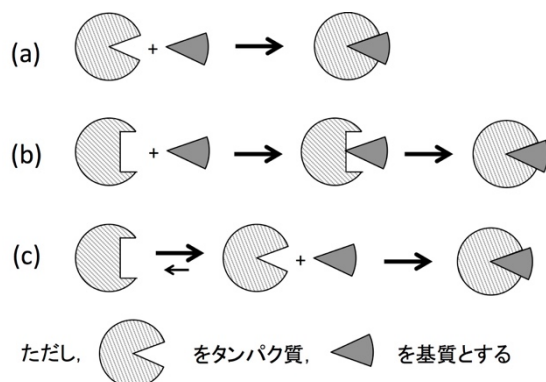
図 32 組の相同なタンパク質ペアについての配列一致度と RMSD.

Chothia, C., and A. M. Lesk. 1986. EMBO J. 5:823-6. を改変

- 1 配列一致度が 30%以上であれば、その構造間の RMSD は、概ね  $1.8\text{\AA}$  以下にとどまる。
- 2 配列一致度が 30%以下になっても、構造間の RMSD は  $2.4\text{\AA}$  以下にとどまる。
- 3 配列一致度が 100%のまったく同じタンパク質どうしであっても、構造どうしが完全に一致するとは限らない。
- 4 配列一致度が 10%程度では構造間の RMSD は小さいが、互いの機能は異なる。

問 57

タンパク質と基質の結合を考える時に代表的な結合モデルとして「鍵と鍵穴 (lock-and-key) モデル」、「induced-fit モデル」、「pre-existing (selected-fit) モデル」が提唱されている。以下の図の(a)~(c)のそれぞれに対応するモデルとして正しい組み合わせを選択肢の中から一つ選べ。



- |   |                      |                      |                      |
|---|----------------------|----------------------|----------------------|
| 1 | (a) lock-and-key モデル | (b) induced-fit モデル  | (c) pre-existing モデル |
| 2 | (a) lock-and-key モデル | (b) pre-existing モデル | (c) induced-fit モデル  |
| 3 | (a) pre-existing モデル | (b) lock-and-key モデル | (c) induced-fit モデル  |
| 4 | (a) induced-fit モデル  | (b) pre-existing モデル | (c) lock-and-key モデル |

問 58

タンパク質分子における他分子との結合サイトとそれを予測する方法についての次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ヘムや ATP など比較的小さな分子が結合するサイトは、くぼんだポケットの形をしていることが多い。
- 2 酵素の活性部位は、特に深いポケットにあることが多い。
- 3 低分子結合サイトの予測では、認識されたポケットにどのような低分子が結合するのかまでは正確にはわからない。
- 4 様々な低分子結合サイトの予測手法が開発されているが、いずれもプローブ球をタンパク質の原子の間に置いて予測する戦略をとっている。

# 問 59

低分子とタンパク質の相互作用について、低分子とタンパク質の各残基が相互作用している時は「1」、相互作用していない時は「0」というプロファイルを考える（相互作用する、しないの定義はここでは問題にしない）。例えば、低分子  $X$  とアミノ酸残基数が  $n$  のタンパク質との相互作用プロファイルは  $\vec{x} = (x_1, x_2, \dots, x_i, \dots, x_n)$  と表現される。このとき  $\vec{x}$  の要素  $x_i$  はタンパク質残基  $i$  と化合物  $x$  との相互作用の有無を意味する。興味のある低分子 - タンパク質複合体における相互作用についてデータベースに存在する複合体情報から得られる類似した相互作用を検索するとき、それらの相互作用プロファイル同士の類似度を Tanimoto 係数 (Jaccard 係数) を用いて計算すると便利である。いま、あるタンパク質と低分子 A と低分子 B の相互作用プロファイルをそれぞれ図に示すようにベクトル  $\vec{a}$ ,  $\vec{b}$  とするとき、これらの類似度を Tanimoto 係数で計算するといくらになるか。選択肢の中から正しいものを一つ選べ。

ただし、各ベクトルの内積について  $\vec{a} \cdot \vec{a} = p$ ,  $\vec{b} \cdot \vec{b} = q$  および  $\vec{a} \cdot \vec{b} = r$  とすると Tanimoto 係数は  $r / (p + q - r)$  で計算される値である。

$$\text{低分子A: } \vec{a} = (00111000101)$$

$$\text{低分子B: } \vec{b} = (00100011010)$$

- 1 1.000
- 2 -0.230
- 3 0.500
- 4 0.125

# 問 60

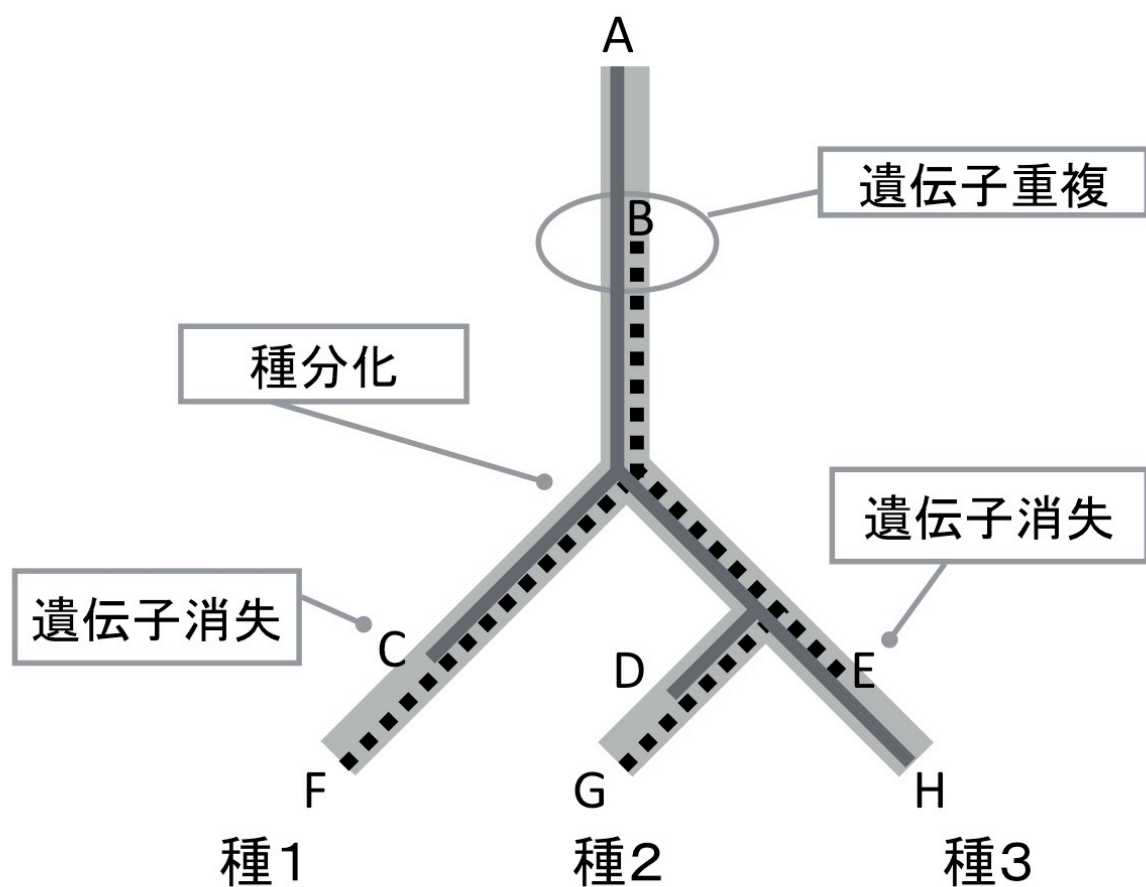
プロテオーム解析技術の発達により、網羅的なタンパク質間相互作用の同定が可能となった。相互作用の有無のみならず、その他の要因も考慮することで、タンパク質自身の機能やタンパク質間相互作用ネットワークの機能について、より詳細な情報を得ることができる。タンパク質間相互作用データを取り扱う際に念頭におくべき事柄についての次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 直接的な相互作用と間接的な相互作用を分けて考える必要がある。
- 2 一過的な相互作用は無視して考える必要がある。
- 3 相互作用相手の多いタンパク質と少ないタンパク質を分けて考える必要がある。
- 4 一度に多数の相手と相互作用するタンパク質と、一度に相互作用する相手の数は少ないが時と場所に応じて相手を変えて相互作用するタンパク質を分けて考える必要がある。

問 61

以下の図は遺伝子 A が遺伝子重複、種分化、および遺伝子消失を経て生物種 1~3 のゲノムに保持される過程を模式的に系統樹として示したものである。遺伝子 B は遺伝子重複時、遺伝子 C、D、E は遺伝子消失時、遺伝子 F、G、H は現存の遺伝子を示す。この分子系統に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 遺伝子 F と G の関係はオーソログである。
- 2 遺伝子 F と遺伝子 H の関係はパラログである。
- 3 遺伝子 F、G、H を用いて作成した系統樹は種系統樹と異なる。
- 4 遺伝子 G と遺伝子 H の分岐時期は遺伝子重複のタイミングではなく種分化のタイミングである。





問 62

遺伝子の進化的性質の 1 つに分子時計がある。分子時計について述べた以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 分子時計とは、アミノ酸置換や塩基置換がほぼ一定の速度で起こる現象を指す。
- 2 分子時計は、ゲノム配列に起こる中立な変異をもとに決定するため、遺伝子によらず同じ速度となる。
- 3 分子時計によって分岐年代を推定するには、化石などの分子以外の証拠から分岐時間がわかっている生物種が一組は必要である。
- 4 分子時計は、非常に離れた生物の間では、同じ塩基に複数回変異が入る可能性があるため精度は低下する。

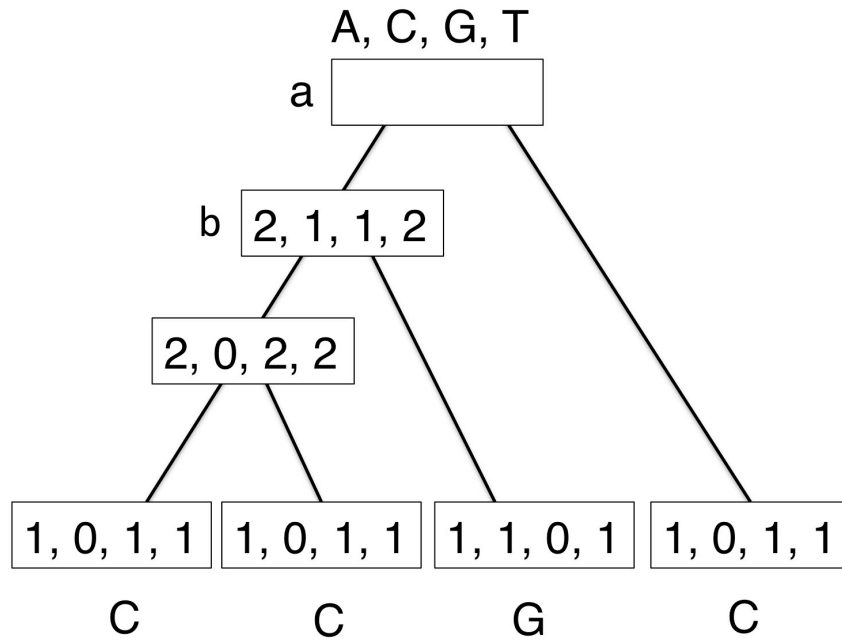
問 63

ある疾患について、十分な規模のゲノムワイド関連解析(GWAS)を行い、 $-\log_{10} p$  値=10 を示す 1 塩基多型(SNP)を発見し、この値が統計学的に有意に高いことを確認した。この結果に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 通常、 $-\log_{10} p$  値の  $p$  は帰無仮説「この SNP は対象疾患に関連しない」を棄却するときの危険率である。
- 2  $-\log_{10} p$  値が 10 であることから、発見された SNP が偽陽性である確率は  $10^{-10}$  である。
- 3  $-\log_{10} p$  値の有意性から、発見された SNP を持つ遺伝子は疾患原因遺伝子であると結論される。
- 4  $-\log_{10} p$  値は、発見された SNP のオッズ比とは厳密には相関しないので、この値から疾患リスクの高さは特定できない。

(問 64・問 65)

以下の図は、ある樹形の系統樹の特定の塩基サイトでの最小置換数を、最大節約法により求めた結果を示している。図中で最下段の C や G は対応する現存遺伝子(OTU)の塩基を、四角内の数字は対応するノードでの塩基がそれぞれ A、C、G、T であった場合の最小置換数を示している。この系統樹に基づいて問 64 および問 65 に答えよ。



問 64

系統樹のノード b の数値の組み合わせを求めるための考え方として、もっとも不適切な記述を選択肢の中から一つ選べ。

- 1 ノード b の塩基が A であれば、左下流のノードで最小値(0)を持つ C に置換(+1)され、右下流の末端ノードの G に置換(+1)される場合が最小置換になるので、値は 2 である。
- 2 ノード b の塩基が C であれば、左下流のノードで最小値(0)を持つ C に置換なし(+0)で受けつがれ、右下流の末端ノードの G に置換(+1)される場合が最小置換になるので、値は 1 である。
- 3 ノード b の塩基が G であれば、左下流のノードで最小値(0)を持つ C に置換(+1)され、右下流の末端ノードの G に置換なし(+0)で受けつがれる場合が最小置換になるので、値は 1 である。
- 4 ノード b の塩基が T であれば、左下流のノードでの最大値(2)と、右下流の末端ノードの最小値(0)の和が最小置換になるので、値は 2 である。

問 65

問 64 をふまえて、系統樹のノード a の空欄に入る数値の組み合わせとして、もっとも適切なものを選択肢の中から一つ選べ。

- 1 3, 1, 1, 3
- 2 3, 1, 2, 3
- 3 3, 2, 1, 3
- 4 3, 2, 2, 3

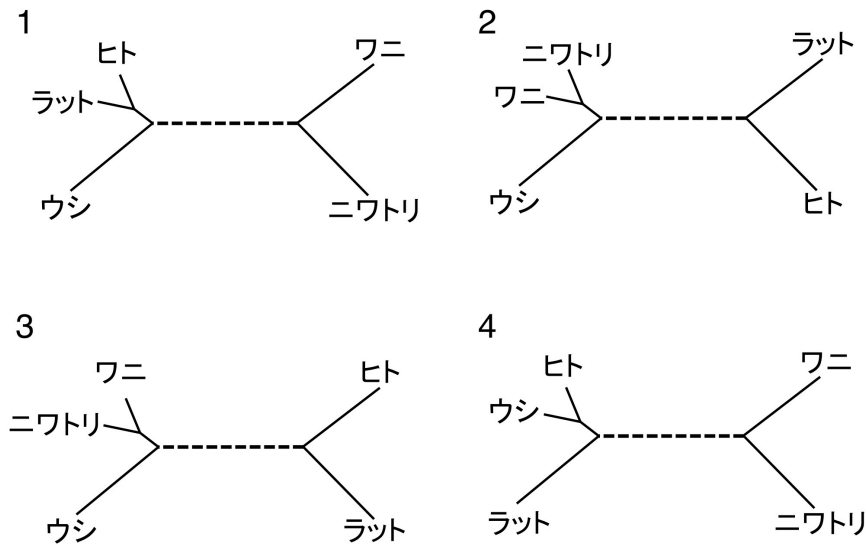
(問 66・問 67)

以下の表は 5 種の生物（ヒト、ラット、ウシ、ニワトリ、ワニ）由来の相同遺伝子について進化距離（置換数/座位）を求めたものである（自明である対角および左下成分は表示していない）。この表に基づいて問 66 および問 67 に答えよ。

|      | ヒト | ラット | ウシ  | ニワトリ | ワニ  |
|------|----|-----|-----|------|-----|
| ヒト   |    | 0.1 | 0.3 | 0.7  | 0.7 |
| ラット  |    |     | 0.3 | 0.7  | 0.7 |
| ウシ   |    |     |     | 0.8  | 0.8 |
| ニワトリ |    |     |     |      | 0.4 |
| ワニ   |    |     |     |      |     |

問 66

これらの進化距離から近隣結合法（NJ 法）で求められる系統樹として、もっとも適切なものを選択肢の中から一つ選べ。系統樹の根（ルート）は省略されており、実線または点線で示された枝の長さ（ブランチ長）は進化距離をほぼ忠実に表しているものとする。なお、点線で示された枝は次の問 67 に関連するが、ここでは実線の枝と区別されないものとする。



問 67

問 66 で求めた系統樹上の、点線で示された枝（ブランチ）の長さとして、もっとも適切な値を選択肢の中から一つ選べ。

- 1 0.2
- 2 0.3
- 3 0.4
- 4 0.5

問 68

ABO 式血液型を決める遺伝子多型は 3 つのアレル A、B、O の持ち方によって決まり、その遺伝子型は AA、AO、BB、BO、AB、OO の 6 通りある。それぞれの遺伝子型を持つとき、血液型の表現型は A、A、B、B、AB、O となる。

今、ハーディ・ワインバーグ平衡にある集団において、血液型の表現型の頻度が A、B、AB、O それぞれ、0.20、0.48、0.16、0.16 のとき、この集団の 3 つのアレル A、B、O のアレル頻度として正しい組み合わせを選択肢の中から一つ選べ。

- |   |   |     |   |     |   |     |
|---|---|-----|---|-----|---|-----|
| 1 | A | 0.1 | B | 0.4 | O | 0.5 |
| 2 | A | 0.2 | B | 0.4 | O | 0.4 |
| 3 | A | 0.3 | B | 0.2 | O | 0.5 |
| 4 | A | 0.3 | B | 0.3 | O | 0.4 |

問 69

次世代シーケンサで解析できる試料として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ヒト血液から抽出したゲノム DNA
- 2 マウスの肝臓 RNA から逆転写反応により合成した cDNA
- 3 免疫沈降法により分離・精製した転写因子
- 4 海水や土壌などから直接 DNA 抽出を行って得た環境 DNA

問 70

次世代シーケンサに関する記述としてもっとも不適切なものを選択肢の中から一つ選べ。

- 1 一度の運転で、数十万から数億にも上る DNA リード数を解読できるため、全ゲノム配列の解読に用いられる。
- 2 解読されてくる DNA リードの数を定量的データとして扱うことで、遺伝子発現量や微生物叢の解析に用いられる。
- 3 次世代シーケンサの機種によっては、DNA のヒストン化学修飾やクロマチン構造を直接解析することができる。
- 4 一度の運転で数十から千以上の数のサンプルを並列解析するために、サンプルを識別するインデックス配列を用いることがある。

# 問 71

次世代シーケンサや DNA マイクロアレイによって、個人のゲノム情報が解析されるようになってきた。個人のゲノム情報に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 性別や属する人類集団、疾患関連変異の有無などが分析できるため、厳重な個人情報管理が必要だと考えられる。
- 2 疾患へのかかりやすさや薬の効きやすさなど個人差の理解が進み、体質に合った医療の実現につながることを期待される。
- 3 認定遺伝カウンセラーと呼ばれる専門家によって、個人ゲノムに関する情報や心理的・社会的サポートが提供される。
- 4 個人のゲノム情報の利用は、保険会社によっても積極的に進められており、保険料の引き下げなどにつながっている。

# 問 72

次世代シーケンサの解析手法と目的の組み合わせとして、もっとも不適切なものを選択肢の中から一つ選べ。

| 選択肢 | 解析手法         | 目的                              |
|-----|--------------|---------------------------------|
| 1   | ChIP-Seq 解析  | 特定のタンパク質と相互作用している DNA 領域の解析     |
| 2   | DNase-Seq 解析 | メチル化されている DNA 塩基の検出             |
| 3   | メタゲノム解析      | 土壌やバイオフィルムなど環境サンプルに含まれる DNA の解析 |
| 4   | RNA-Seq 解析   | RNA から逆転写した cDNA の配列と発現量の解析     |

# 問 73

次世代シーケンサのデータは、FASTQ という形式で出力されることが多い。以下のうち FASTQ ファイルに関する記述として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 FASTQ はバイナリファイルなので、そのままでは閲覧することができない。
- 2 FASTQ は、塩基配列と各塩基のクオリティスコアを一緒に記録している。
- 3 FASTQ は、FastQC などの様々なクオリティ評価ソフトウェアで分析できる。
- 4 FASTQ は、そのまま BWA などを用いたマッピング処理に使用できる。

#### 問 74

がん細胞と非がん細胞に対してマイクロアレイの発現量解析を行った結果、がん細胞で有意に発現が異なる 30 遺伝子を検出する事ができた。このマイクロアレイの結果の情報解析を行う操作としてもっとも不適切なものを選択肢の中から一つ選べ。

- 1    がんで多く見られた 30 遺伝子には、該当遺伝子上に必ず変異部位があるため、この中に含まれる一塩基多型(SNV)を探しそれががんの原因であると推定する。
- 2    この 30 遺伝子を DAVID 等のツールを用いることで、有意に濃縮（エンリッチ）している Gene Ontology term を検索し、遺伝子機能を推定する。
- 3    この 30 遺伝子が有意に濃縮しているパスウェイを探すことで、どの代謝経路ががんに関係しているのかを推定する。
- 4    30 遺伝子の転写制御領域にどのようなモチーフが濃縮しているかを調べることで、転写因子結合部位と制御関係を推定する。

#### 問 75

オーミクス解析の説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1    ゲノム解析は、次世代シーケンサや SNP アレイなどを用いて、生物の DNA 配列の全長や個人差などを調べる手法である。
- 2    トランスクリプトーム解析は、サンプル内の RNA 量に着目し、どのような転写産物があるかを推定する手法である。
- 3    プロテオーム解析は、サンプルに含まれるタンパク質の立体構造を網羅的に決定する手法である。
- 4    メタボローム解析は、質量分析計や NMR などを用い、サンプル内の代謝産物に着目する技術である。

問 76

5つのタンパク質（タンパク質 A、B、C、D、E）の相互作用を酵母ツーハイブリッド法を用いて観察した結果、以下の相互作用が観察された。このときもっとも不適切なものを選択肢の中から一つ選べ。

タンパク質 A – タンパク質 B

タンパク質 B – タンパク質 B

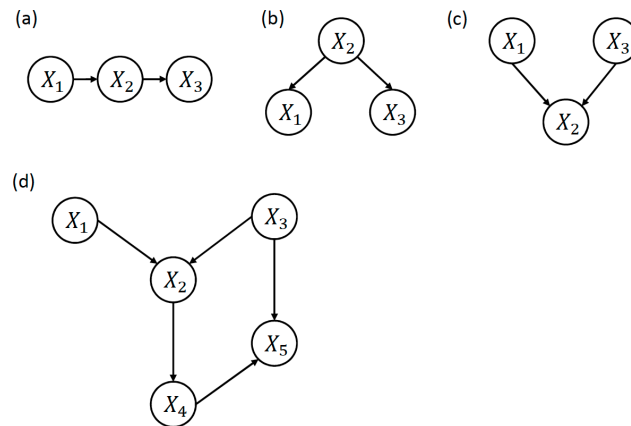
タンパク質 A – タンパク質 C

タンパク質 D – タンパク質 E

- 1 タンパク質 B とタンパク質 C は、直接結合する可能性がある。
- 2 タンパク質 B は、ホモダイマーを形成する可能性がある。
- 3 タンパク質 A、B、C は、複合体を形成している可能性がある。
- 4 タンパク質 D とタンパク質 E は、直接結合する可能性がある。

(問 77・問 78)

遺伝子制御ネットワーク推定で用いられるベイジアンネットワークは確率変数間の条件付独立性を表す有向非巡回グラフである。ベイジアンネットワークにおいて 3 つの確率変数からなるグラフの条件付独立性は、図(a)、(b)、(c)の 3 つの場合がある。 $X_2$ で条件付けた場合、(ア)において $X_1$ と $X_3$ は独立となるが(イ)において独立とならない。逆に $X_2$ で条件付けない場合、前記と逆の関係となる。上記の 3 つの場合において $X_2$ の状態によって $X_1$ と $X_3$ が確率的に独立となる場合、 $X_1$ と $X_3$ の間のパスは $X_2$ においてブロックされていると定義する。より多数の確率変数からなるベイジアンネットワークで考えた場合についても、2 つの確率変数間を結ぶパスの全てがブロックされる場合のみ、対象の 2 つの確率変数は条件付独立となることが知られている。図(d)の場合、 $X_1$ と $X_5$ の間にはパス $X_1-X_2-X_3-X_5$ とパス $X_1-X_2-X_4-X_5$ の 2 つのパスが存在するが、 $X_4$ で条件付けた場合、パス $X_1-X_2-X_3-X_5$ では(ウ)において、パス $X_1-X_2-X_4-X_5$ では(エ)においてブロックされるため、 $X_1$ と $X_5$ は、 $X_4$ が与えられた下で条件付独立である。



以下の問 77 および問 78 に答えよ。

問 77

(ア) と (イ) に入る記述としてもっとも適切なものを選択肢の中から一つ選べ。

- |   |             |             |
|---|-------------|-------------|
| 1 | (ア) 図 a と c | (イ) 図 b     |
| 2 | (ア) 図 a と b | (イ) 図 c     |
| 3 | (ア) 図 a     | (イ) 図 b と c |
| 4 | (ア) 図 c     | (イ) 図 a と b |

問 78

(ウ) と (エ) に入る記述としてもっとも適切なものを選択肢の中から一つ選べ。

- |   |           |           |
|---|-----------|-----------|
| 1 | (ウ) $X_2$ | (エ) $X_4$ |
| 2 | (ウ) $X_2$ | (エ) $X_2$ |
| 3 | (ウ) $X_3$ | (エ) $X_2$ |
| 4 | (ウ) $X_3$ | (エ) $X_4$ |



(問 79・問 80)

二次元の時間の関数ベクトル $\mathbf{y}(t)$ からなる連立微分方程式を下記の式で表す。

$$\frac{d}{dt}\mathbf{y}(t) = \mathbf{b} - A\mathbf{y}(t).$$

$\mathbf{y}(t)$ の初期値を $\mathbf{y}(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ とし、 $A$ と $\mathbf{b}$ はそれぞれ下記の $2 \times 2$  行列と長さ 2 のベクトルで与えられるものとする。

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

$A$ は固有値 $\lambda_1$ と $\lambda_2$ を用いて下記のように対角化可能である。

$$A = P^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P$$

この時、 $\lambda_1$ と $\lambda_2$ はそれぞれ（ ア ）となる。 $\mathbf{y}(t)$ の一般解は 0 でない定数 $C_1$ 、 $C_2$ を用いて下記で与えられる。

$$\mathbf{y}(t) = A^{-1}\mathbf{b} + P^{-1} \begin{pmatrix} C_1 \exp(-\lambda_1 t) \\ C_2 \exp(-\lambda_2 t) \end{pmatrix},$$

$t$ の値が十分大きくなるにつれて  $\mathbf{y}(t)$ は（ イ ）する。

以下の問 79 および問 80 に答えよ。

問 79

（ ア ）に入る記述としてもっとも適当なものを選択肢の中から一つ選べ。

- 1    -5 と 2
- 2    2 と 5
- 3    3 と 4
- 4    -4 と -3

問 80

（ イ ）に入る記述としてもっとも適当なものを選択肢の中から一つ選べ。

- 1    双方の固有値が正であることから $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ に収束
- 2    双方の固有値が正であることから $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ に収束
- 3    片方の固有値が負であることから発散
- 4    双方の固有値が負であることから $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ に収束

平成 28 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  
【解答】

|      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|
| 問 1  | 問 2  | 問 3  | 問 4  | 問 5  | 問 6  | 問 7  | 問 8  | 問 9  | 問 10 |
| 4    | 2    | 1    | 4    | 2    | 3    | 4    | 3    | 4    | 3    |
| 問 11 | 問 12 | 問 13 | 問 14 | 問 15 | 問 16 | 問 17 | 問 18 | 問 19 | 問 20 |
| 2    | 2    | 1    | 1    | 4    | 3    | 2    | 1    | 3    | 4    |
| 問 21 | 問 22 | 問 23 | 問 24 | 問 25 | 問 26 | 問 27 | 問 28 | 問 29 | 問 30 |
| 2    | 3    | 4    | 1    | 3    | 1    | 1    | 3    | 1    | 4    |
| 問 31 | 問 32 | 問 33 | 問 34 | 問 35 | 問 36 | 問 37 | 問 38 | 問 39 | 問 40 |
| 3    | 2    | 3    | 1    | 2    | 4    | 1    | 1    | 3    | 1    |
| 問 41 | 問 42 | 問 43 | 問 44 | 問 45 | 問 46 | 問 47 | 問 48 | 問 49 | 問 50 |
| 4    | 3    | 2    | 4    | 3    | 1    | 1    | 2    | 3    | 1    |
| 問 51 | 問 52 | 問 53 | 問 54 | 問 55 | 問 56 | 問 57 | 問 58 | 問 59 | 問 60 |
| 4    | 3    | 2    | 3    | 4    | 4    | 1    | 4    | 4    | 2    |
| 問 61 | 問 62 | 問 63 | 問 64 | 問 65 | 問 66 | 問 67 | 問 68 | 問 69 | 問 70 |
| 4    | 2    | 3    | 4    | 2    | 1    | 3    | 2    | 3    | 3    |
| 問 71 | 問 72 | 問 73 | 問 74 | 問 75 | 問 76 | 問 77 | 問 78 | 問 79 | 問 80 |
| 4    | 2    | 1    | 1    | 3    | 1    | 2    | 1    | 2    | 2    |

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。