

**平成30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

真核生物の遺伝情報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 ヒトの体細胞のミトコンドリアには DNA が含まれており、個体の性別に関わらず、その DNA にコードされた遺伝情報は主に母親から伝わる。
- 2 オスのマウスの体細胞が持つ X 染色体は主に父親から伝わる。
- 3 キイロシヨウジョウバエの体細胞のミトコンドリアに局在するタンパク質の中には、その遺伝情報が核の DNA にコードされたものがある。
- 4 シロイヌナズナの体細胞の葉緑体には DNA が含まれており、その DNA にコードされた遺伝情報は進化の過程でシアノバクテリア（藍藻、藍色細菌）の一種が細胞内共生したことに由来すると考えられている。

### 問 2

セントラルドグマ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 DNA の遺伝情報は、DNA ポリメラーゼの働きにより mRNA に転写される。
- 2 転写された真核細胞の mRNA は、核から細胞質へ移動し、細胞質中のリボソームと結合する。
- 3 アミノ酸と tRNA が共有結合したアミノアシル tRNA が、mRNA 上の対応するコドンと対合することによって、mRNA の遺伝情報はタンパク質に翻訳される。
- 4 ウイルスの中には、宿主細胞への感染後に自身の RNA を DNA に逆転写するものもある。

### 問 3

ウイルス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 ウイルスは自己増殖することができず、ウイルスの遺伝情報をコードする核酸（DNA または RNA）は宿主の細胞内で複製される。
- 2 真正細菌（バクテリア）を宿主とするウイルスはバクテリオファージあるいはファージと呼ばれる。
- 3 ウイルスのカプシドと呼ばれるタンパク質は宿主の細胞の翻訳機構によって生合成される。
- 4 脊椎動物には様々な生体防御のしくみが備わっているため、外胚葉由来の組織の細胞にウイルスが感染することはあっても、中胚葉由来の組織の細胞にウイルスが感染することはない。

#### 問 4

減数分裂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 動物の生殖細胞である卵や精子は減数分裂を通じて形成される。
- 2 減数分裂により形成される生殖細胞と体細胞の染色体の数は同じである。
- 3 1 個の卵母細胞から 2 回の細胞分裂を経て 1 個の卵細胞が形成される。
- 4 減数分裂時に相同組み換えが起こり、配偶子の遺伝的多様性が生じる。

#### 問 5

核酸の生合成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。なお、文中で用いられている DNA 依存性とは DNA を鋳型に使用するという意味、RNA 依存性とは RNA を鋳型に使用するという意味である。

- 1 原核生物の細胞では、一般に 3 種類の DNA 依存性 RNA ポリメラーゼ (I, II, III) がそれぞれ rRNA、mRNA、tRNA の生合成 (転写) を行う。
- 2 真核生物の核における DNA 複製では、一般に 2 種類の DNA 依存性 DNA ポリメラーゼ ( $\delta, \epsilon$ ) が主な役割を担っている。
- 3 cDNA の合成に用いられる逆転写酵素は RNA 依存性 DNA ポリメラーゼの一種である。
- 4 ポリオウイルスは+鎖 RNA ウイルスで、ウイルスゲノムの複製には自身がコードする RNA 依存性 RNA ポリメラーゼが用いられる。

#### 問 6

エピジェネティクスに関する以下の記述について ( a ) から ( c ) に入る語句の組み合わせとして適切なものはどれか。選択肢の中から一つ選べ。

エピジェネティクスとは DNA 塩基配列の変化を ( a )、細胞分裂後も継承される遺伝子発現あるいは細胞表現型の変化のことで、主に ( b ) の ( c ) による。

- |           |              |         |
|-----------|--------------|---------|
| 1 (a)伴う   | (b)DNA       | (c)塩基置換 |
| 2 (a)伴わない | (b)DNA やヒストン | (c)化学修飾 |
| 3 (a)伴わない | (b)テロメア      | (c)短縮   |
| 4 (a)伴う   | (b)アミノ酸側鎖    | (c)化学修飾 |

問 7

免疫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 細菌などの異物が体内に侵入したときに補体や食細胞などにより対応する仕組みを自然免疫という。
- 2 抗原刺激により形質細胞に分化した 1 つの B 細胞から、多種類の抗体が産生される。
- 3 MHC 分子は免疫応答において自己と非自己を識別して異物を認識する役目を持っている。
- 4 生体が体内に侵入した異物を認識、学習してリンパ球などにより対応する仕組みを獲得免疫という。

問 8

シグナル伝達に関する以下の記述について、( a ) から ( c ) に入る語句の組み合わせとして適切なものはどれか。選択肢の中から一つ選べ。

真核生物の細胞間シグナル伝達においては ( a ) が細胞膜上の ( b ) に結合することをきっかけに、細胞内のタンパク質の構造や酵素の活性に変化が生じて、細胞の応答につながる経路と、核における ( c ) の制御を通じて細胞の応答につながる経路があり、一般的に前者のほうが細胞はすばやく応答できる。

- |                |              |            |
|----------------|--------------|------------|
| 1 (a)抗原        | (b)抗体        | (c)タンパク質合成 |
| 2 (a)細胞外シグナル因子 | (b)受容体タンパク質  | (c)遺伝子発現   |
| 3 (a)受容体タンパク質  | (b)細胞外シグナル因子 | (c)タンパク質合成 |
| 4 (a)転写因子      | (b)プロモーター    | (c)遺伝子発現   |

問 9

PCR 法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 RNA を PCR 反応の鋳型に直接使用することが可能である。
- 2 反応には耐熱性の DNA ポリメラーゼが利用される。
- 3 PCR 反応の鋳型として 1 本鎖 DNA を使用することが可能である。
- 4 DNA の合成の開始にはプライマーが必要である。

問 10

クローニングとシーケンス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 大腸菌と酵母のように異なる生物種の両方で増殖するプラスミドベクターは、おのおのの生物種で機能する異なる複数の複製開始点を持つ。
- 2 すでに解読された生物種のゲノム配列に基づいて、異なる個体や近縁の種のゲノム配列を解析する方法を *de novo* (デノボ) アセンブリという。
- 3 多数の DNA クローンの塩基配列の重なりを見つけながら、それらのクローンをアセンブルすることによってゲノム全体の配列を知ることができる。
- 4 全ゲノムショットガン法の手法の 1 つとして、超音波処理によりゲノム DNA を断片化し、その断片をベクターに組み込む方法が用いられる。

問 11

iPS 細胞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 体中の様々な細胞に変化する能力をもつ多能性細胞である。
- 2 初期化因子の導入にレトロウイルスベクターを用いることや未分化細胞の混入などから、腫瘍化の危険性がないとは言い切れない。
- 3 皮膚や血液など採取しやすい体細胞を用いて、人為的に作製が可能である。
- 4 患者自身の細胞から作製した場合、移植時に拒絶される可能性が高い。

問 12

ゲノム編集に関する以下の記述について ( 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はどれか。選択肢の中から一つ選べ。

ゲノム編集は ( a ) 配列上の任意の場所に欠失、置換、挿入を導入する技術であり、部位特異的な ( b ) を用いる。現在は CRISPR-Cas9 法が主流となっているが、標的外の切断が生じる ( c ) が問題になることがある。

- |            |           |              |
|------------|-----------|--------------|
| 1 (a)ゲノム   | (b)ヌクレアーゼ | (c)オフターゲット効果 |
| 2 (a)ゲノム   | (b)プロテアーゼ | (c)ワールブルグ効果  |
| 3 (a)タンパク質 | (b)ヌクレアーゼ | (c)オフターゲット効果 |
| 4 (a)タンパク質 | (b)プロテアーゼ | (c)ワールブルグ効果  |

問 13

ヒトの視細胞（光受容細胞）におけるシグナル伝達に関して順に説明し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 ヒトの網膜には桿体細胞と錐体細胞の 2 種類の視細胞があり、明暗視に関与する光受容体であるロドプシンは桿体細胞で発現する。
- 2 視細胞の光受容体が光を受容すると、会合しているトランスデューシン（ヘテロ三量体 G タンパク質の一種）が活性化される。
- 3 続いて、制限酵素が活性化され、セカンドメッセンジャーである cGMP（サイクリック GMP）が分解される。
- 4 視細胞内の cGMP 濃度が低下すると、cGMP 依存性陽イオンチャネルが閉じ、細胞の膜電位が変化する。

問 14

ATP 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 動物の細胞において、受容体型チロシンキナーゼは基質となるタンパク質の特定のチロシン残基に ATP のリン酸基を転移（リン酸化）することによって細胞外のシグナルを細胞内に伝達する。
- 2 真核生物のミトコンドリアには電子伝達系の様々な酵素が局在しており、これら代謝系によって生じたプロトン濃度勾配を利用して、ATP 合成酵素により ATP が生合成される。
- 3 真核生物のミトコンドリアにおいて DNA が複製される際には、CTP、GTP、TTP とともに ATP も DNA 合成の基質として用いられる。
- 4 細胞の ATP 消費量に応じてミトコンドリアの数は異なるが、ATP 供給が必要な筋肉細胞や神経細胞にはとりわけ多く存在する。

問 15

原核生物および真核生物の多くの種で用いられる遺伝暗号のことを普遍遺伝暗号とよぶ。遺伝暗号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 普遍遺伝暗号には、アルギニン、セリンおよびロイシンを指定するコドンがそれぞれ 6 つある。
- 2 普遍遺伝暗号には、イソロイシンを指定するコドンが 3 つある。
- 3 普遍遺伝暗号には、翻訳終了を意味する終止コドンが 2 つある。
- 4 ヒトのミトコンドリアで用いられる遺伝暗号は、普遍遺伝暗号と一部が異なっている。

問 16

次の[ ]内に示したアミノ酸の“名称、3文字表記、1文字表記”のうち、不適切な表記を含むものを選択肢の中から一つ選べ。

- 1 [アスパラギン、Asn、N]
- 2 [グルタミン酸、Glu、E]
- 3 [トリプトファン、Trp、Y]
- 4 [リジン、Lys、K]

問 17

ゲノム配列の多型を表す用語として不適切なものを選択肢の中から一つ選べ。

- 1 SNP
- 2 CNV
- 3 VNTR
- 4 GWAS

問 18

ヒトゲノム・遺伝子解析研究に関する倫理指針（平成 29 年 2 月 28 日一部改正）において定められている、全ての研究者等の基本的な責務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 全ての研究者等は、ヒトゲノム・遺伝子解析研究の社会的有益性を確認するとともに、個人の人權の保障を科学的又は社会的な利益に優先して配慮しなければならない。
- 2 全ての研究者等は、提供者又は代諾者等から同意を得られない場合も科学的な利益に配慮し、ヒトゲノム・遺伝子解析研究を実施することが許容される。
- 3 全ての研究者等は、個人情報の保護を図るとともに、個人情報の取扱いに関する苦情等に誠実に対応しなければならない。
- 4 全ての研究者等は、ヒトゲノム・遺伝子解析研究の実施に当たっては、偽りその他不正の手段により個人情報及び試料・情報を取得してはならない。

問 19

ゲノム情報は特定の個人を識別しうる情報であることから、その取扱いは個人情報保護法の対象となりうる。「個人情報の保護に関する法律についてのガイドライン・通則編」(平成 29 年 3 月一部改正)によると、特定の個人を識別することのできる「個人識別符号」に該当するとみなされるデータとして、もっとも不適切なものはどれか。選択肢の中から一つ選べ。

- 1 全エクソームシーケンスデータ
- 2 全ゲノム塩基多型(SNP)データ
- 3 互いに独立な 4 箇所の SNP から構成されるシーケンスデータ
- 4 9 座位以上の 4 塩基単位の繰り返し配列(short tandem repeat; STR)に基づく遺伝型情報

問 20

タンパク質の翻訳後修飾として不適切なものを選択肢の中から一つ選べ。

- 1 ユビキチン化
- 2 メチル化
- 3 グリコシル化
- 4 ポリアデニル化

問 21

0.0 以上 1.0 以下の浮動小数点数 $x$ に対して以下の 4 つの計算を行う。この中で、計算結果が他の 3 つと異なる可能性があるものを選択肢の中から一つ選べ。ただし、 $\ll 1$  は C 言語の仕様に即し左に 1 ビットシフトを行う演算とし、 $*$ は乗算演算子であるものとする。また、`int` 型への型変換は C 言語の仕様に即し、小数点以下が 0 に丸められる(切り捨てられる)ものとする。

- 1  $(\text{int})x + (\text{int})x$
- 2  $2 * (\text{int})x$
- 3  $((\text{int})x) \ll 1$
- 4  $(\text{int})(x + x)$

問 22

コンピュータ上でのデータの記述形式には、様々なものが開発されているが、その中でもXML(Extensible Markup Language)はもっとも普及した形式の一つである。次に示した説明文の中で、XML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 XMLは、統計処理に特化したプログラミング言語の一種である。
- 2 XMLは、バイナリデータではなく、テキストデータとして保存される。
- 3 XMLでは、各要素をタグで囲み、階層的なデータ構造を入れ子で表現している。
- 4 XMLでは、DTD(Document Type Definition)によって文書構造を定義することができる。

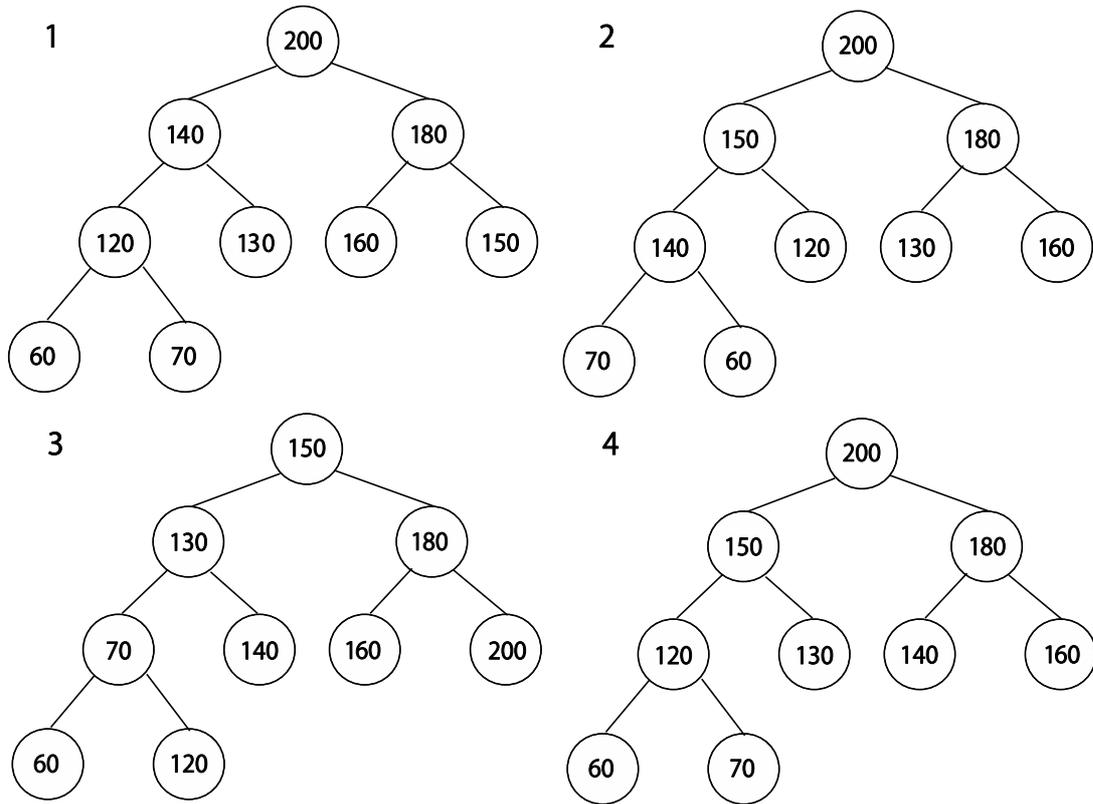
問 23

長さ  $n$  のソートされた整数配列上の二分探索に関する以下の記述のうち、不適切なものを選択肢の中から一つ選べ。

- 1 配列中に整数  $a$  が存在するかどうかを  $O(\log n)$  時間で確かめることが可能である。
- 2 配列中に 2 つの整数  $a$  および  $b$  がいずれも存在するか否かを  $O(\log n)$  時間で確かめることが可能である。
- 3 配列中に存在する  $a$  以上  $b$  以下の整数の数を  $O(\log n)$  時間で計算することが可能である。
- 4 配列中に存在する  $a$  以上  $b$  以下の整数を  $O(\log n)$  時間ですべて列挙することが可能である。

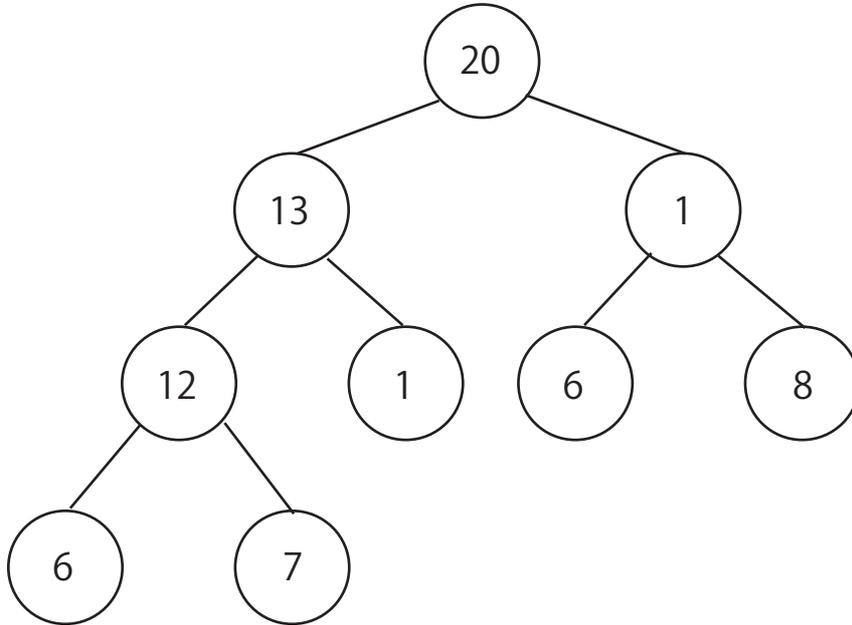
問 24

次の4つの木は、同じ数字集合を木構造のノードに割り当てたものである。4つのうち3つがヒープ構造になっており、残りの一つが二分探索木になっている。二分探索木となっているものを4つの木から一つ選べ。



問 25

下の二分木を先行順(pre-order)で探索した場合のラベル列として正しいものを選択肢の中から一つ選べ。



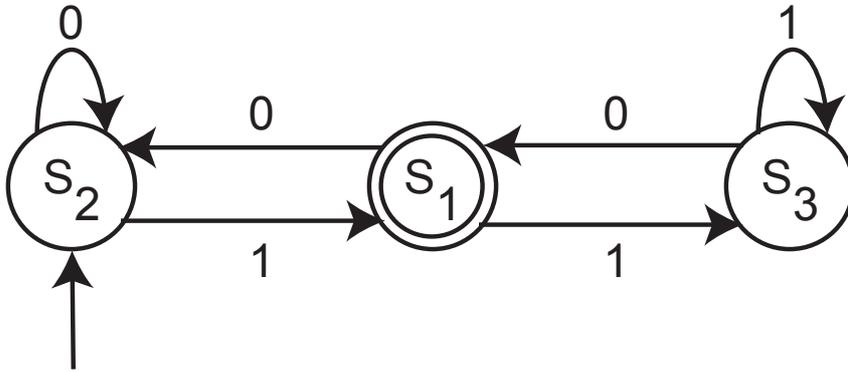
なお、参考までに、以下に、木のルートを入力として先行順を出力する擬似コードを示す。ただし、「foreach ( $w \in \text{children}(v)$ ) do ...」は、ノード  $v$  の子のノードそれぞれに対し左側の子のノードから順番にアクセスすることを表すものとする。

```
PreOrder (node v):  
begin  
  output (v のラベル);  
  foreach (c ∈ children(v)) do begin  
    PreOrder (c);  
  end  
end
```

- 1 6, 7, 12, 1, 13, 6, 8, 1, 20
- 2 6, 12, 7, 13, 1, 20, 6, 1, 8
- 3 20, 13, 1, 12, 1, 6, 8, 6, 7
- 4 20, 13, 12, 6, 7, 1, 1, 6, 8

問 26

下の遷移図は  $S_1, S_2, S_3$  の 3 つの状態をもつ有限オートマトンである。 $S_2$  を初期状態、 $S_1$  を受理状態とする。この有限オートマトンが受理する記号列を選択肢の中から一つ選べ。

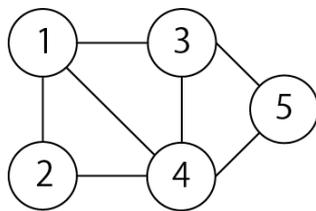


- 1 10000011
- 2 11111100
- 3 11000011
- 4 11111110

問 27

グラフのデータ構造と計算量に関する以下の図および説明文において、( a ), ( b ), ( c ), ( d )に入る数字および語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。

グラフは、有限個の頂点からなる集合と、2 頂点を接続するいくつかの辺からなる集合によって定義される。グラフを計算機に格納するためのデータ構造の一例として、次に定義する隣接行列が挙げられる。隣接行列は行数・列数がそれぞれグラフの頂点数と等しい正方行列で、その $(i, j)$ 成分は、頂点  $i$  と頂点  $j$  を結ぶ辺の数で定義される。例えば、左下図のような 1~5 を頂点とするグラフが与えられたとき、対応する隣接行列は右下図のようになる。グラフの頂点の個数を  $n$ , 辺の本数を  $m$  とするとき、隣接行列の空間計算量は( a )であり、任意の 2 頂点間に辺が存在するか否かを時間計算量 ( b )で判定することができる。



	1	2	3	4	5
1	0	1	1	1	0
2	(c)				
3	1	0	0	1	1
4	(d)				
5	0	0	1	1	0

- |   |              |            |               |               |
|---|--------------|------------|---------------|---------------|
| 1 | (a) $O(m+n)$ | (b) $O(1)$ | (c) 1 1 1 0 1 | (d) 1 0 0 1 0 |
| 2 | (a) $O(m+n)$ | (b) $O(m)$ | (c) 1 1 1 0 1 | (d) 1 0 0 1 0 |
| 3 | (a) $O(n^2)$ | (b) $O(1)$ | (c) 1 0 0 1 0 | (d) 1 1 1 0 1 |
| 4 | (a) $O(n^2)$ | (b) $O(m)$ | (c) 1 0 0 1 0 | (d) 1 1 1 0 1 |

問 28

次の RNA の二次構造とその予測法に関する記述として不適切なものを選択肢の中から一つ選べ。

- 1 RNA の二次構造は、互いに相補的な配列が塩基対を形成することによりできるステムと、それらのあいだで 1 本鎖として存在するループで構成される。
- 2 RNA の二次構造においては、通常の G-C、A-U (ワトソン・クリック対) の他に、G-U 対も塩基対を形成する。
- 3 RNA の二次構造はシュードノット (偽結び目構造) と呼ばれる構造まで含めて文脈自由文法で表現できる。
- 4 Nussinov アルゴリズムはシュードノットを含まない RNA の二次構造を予測することができる。

問 29

化合物の標準的な表現にフィンガープリントと呼ばれるバイナリベクトル (0 または 1 を要素とするベクトル) 表現がある。なお、以下では、フィンガープリントの要素がすべて 0 となることはないものとし、また、ベクトル、整数列等のインデックスはすべて 0 から始めるものとする。また、以下では、 $\wedge$ ,  $\vee$  はそれぞれ、フィンガープリントの各次元ごとの論理積、フィンガープリントの各次元ごとの論理和を表し、 $|X|$  はフィンガープリント  $X$  中の 1 の個数、 $X[i]$  はフィンガープリント  $X$  の  $i$  番目の値を表す。たとえば、 $A=(1, 0, 0, 1, 1)$ 、 $B=(0, 1, 0, 1, 1)$  に対し、 $|A \wedge B|=2$ 、 $|A \vee B|=4$ 、 $|A|=3$ 、 $A[3]=1$  である。

2 つのフィンガープリント間の類似度として Tanimoto 類似度が標準的に使われる。2 つのフィンガープリント  $A, B$  に対する Tanimoto 類似度は、 $J(A, B) = |A \wedge B| / |A \vee B|$  として定義される。

フィンガープリント  $X$  は  $X[i]=1$  であるような整数  $i$  を昇順に並べた列としても表現でき、これをフィンガープリント  $X$  の整数列表現とよぶものとする。例えば、フィンガープリント  $A=(1, 0, 0, 1, 1)$  に対する整数列表現は  $V=\{0, 3, 4\}$  である。下の擬似コードは整数列表現による 2 つのフィンガープリント  $V, U$  の Tanimoto 類似度を計算する擬似コードである。 $V[i], U[j]$  は整数列表現  $V$  の  $i$  番目の値、整数列表現  $U$  の  $j$  番目の値を表す。 $V.len$  と  $U.len$  は  $V$  と  $U$  の長さを表す。 $V[i], U[i], V.len, U.len$  は  $O(1)$  時間で計算できるものとする。

```

1.  CompTanimoto(V, U) :
2.  begin
3.      sum ← 0 ;
4.      i ← 0 ;
5.      j ← 0 ;
6.      while ((i < V.len) かつ (j < U.len)) {
7.          if (V[i] < U[j])
8.              { i ← i+1 ; }
9.          else if (V[i] > U[j])
10.             { j ← j+1 ; }
11.          else {
12.              sum ← sum + 1 ;
13.              i ← i+1 ;
14.              j ← j+1 ;
15.          }
16.      }
17.      return sum / (V.len + U.len - sum) ;
18. end

```

擬似コードの計算時間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  $O(V.len + U.len)$
- 2  $O(\log(V.len + U.len))$
- 3  $O(V.len \times U.len)$
- 4  $O(\log(V.len \times U.len))$

問 30

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件数として適切なものを選択肢の中から一つ選べ。

```
SELECT ID FROM 健康診断結果
WHERE 睡眠時間 < 6.0 AND 体重 > 60.0;
```

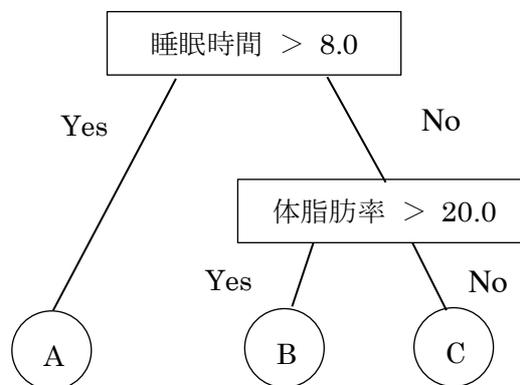
表：健康診断結果

ID	身長	体重	体脂肪率	睡眠時間	日常的な運動
APF199R77	173.4	66.3	9.3	7.5	している
UCF329B13	169.9	66.7	23.0	5.5	していない
RRD910X00	153.3	49.8	20.5	8.5	している
DRZ353B36	187.0	78.4	11.9	5.5	している
BBF001Y21	166.9	55.3	17.1	5.5	している
OPV214J66	161.6	82.5	33.1	4.5	していない

- 1 1件
- 2 2件
- 3 3件
- 4 4件

問 31

下の図は、問 30 の表において「睡眠時間」と「体脂肪率」から「日常運動」の項目を推定する決定木の一例である。以下の選択肢の中から、問 30 の表のデータにこの決定木を適用した時に正答数が最も多くなるようなノード A,B,C に入れる言葉の組み合わせとして、適切なものを選択肢の中から一つ選べ。



- |   |          |          |          |
|---|----------|----------|----------|
| 1 | A: している  | B: している  | C: している  |
| 2 | A: している  | B: していない | C: している  |
| 3 | A: している  | B: していない | C: していない |
| 4 | A: していない | B: している  | C: していない |

問 32

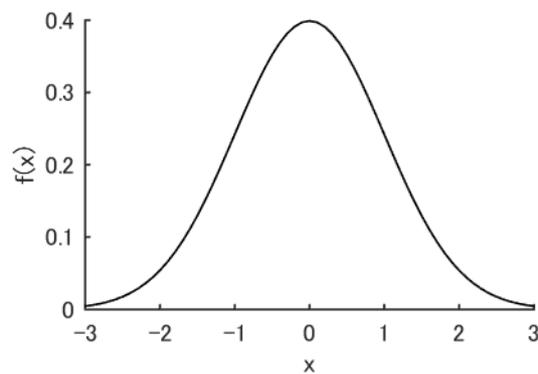
確率分布に関する以下の説明文において、( a ) と ( b ) に入る語の組み合わせとして正しいものを選択肢の中から一つ選べ。

2 種類の可能な結果 ( 成功と失敗など ) を生じる試行をベルヌーイ試行という。ポアソン分布は、同じ条件で独立に複数回ベルヌーイ試行を繰り返したときに現れる確率分布である ( a ) において、試行回数が非常に多く、成功確率が非常に小さいときの極限をとった分布とみなすことができる。また、ポアソン分布においては期待値と ( b ) が等しくなることが大きな特徴である。

- |   |          |        |
|---|----------|--------|
| 1 | (a)2 項分布 | (b)中央値 |
| 2 | (a)2 項分布 | (b)分散  |
| 3 | (a)正規分布  | (b)中央値 |
| 4 | (a)正規分布  | (b)分散  |

問 33

$f(x)$  を標準正規分布の確率密度関数とする。  $0 < a < b < +\infty$  なる実数  $a, b$  に対して、もっとも適切なものを選択肢の中から一つ選べ。なお、以下の図は  $f(x)$  をプロットしたものである。



- |   |                   |
|---|-------------------|
| 1 | $f(a) < 0 < f(b)$ |
| 2 | $f(b) < 0 < f(a)$ |
| 3 | $0 < f(a) < f(b)$ |
| 4 | $0 < f(b) < f(a)$ |

問 34

$F(x)$  を標準正規分布の累積密度関数とする。 $-\infty < a < 0 < b < +\infty$  なる実数  $a, b$  に対して、もっとも適切なものを選択肢の中から一つ選べ。なお、 $F(x)$  は、標準正規分布の確率密度関数  $f(t)$  に対し  $F(x) := \int_{-\infty}^x f(t)dt$  と定義される。

- 1  $F(a) < 0 < F(b)$
- 2  $F(b) < 0 < F(a)$
- 3  $0 < F(a) < F(b)$
- 4  $0 < F(b) < F(a)$

問 35

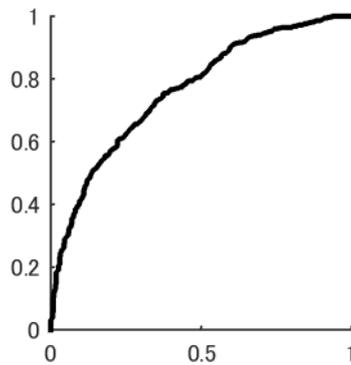
事象  $A$  および  $B$  に対して、それらの排反事象をそれぞれ  $\bar{A}$  および  $\bar{B}$  で表すことにする。条件付確率  $P(A|B)$  に関して、もっとも適切な等式を選択肢の中から一つ選べ。

- 1  $P(A|B) = 1 - P(\bar{A}|\bar{B})$
- 2  $P(A|B) = 1 - P(\bar{A}|B)$
- 3  $P(A|B) = 1 - P(A|\bar{B})$
- 4  $P(A|B) = 1 - P(A|B)$

問 36

2 値分類問題に対する予測手法の評価に ROC 曲線がしばしば用いられる。以下の ( a ) ~ ( d ) 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。

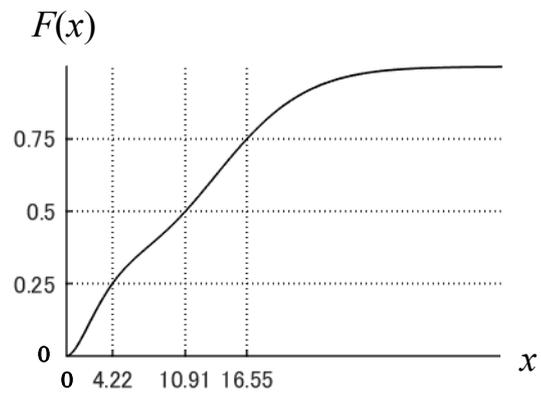
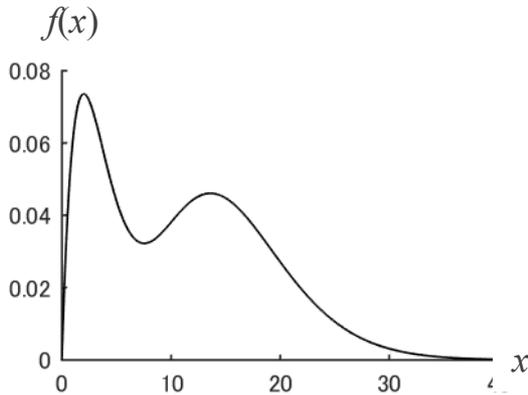
図のような ROC 曲線を考える。ROC 曲線の横軸は (1-特異度) を表し、縦軸は ( a ) を表す。TP, FP, FN, TN をそれぞれ真陽性数, 偽陽性数, 偽陰性数, 真陰性数を表すとすると, 特異度は ( b ) のように定義され、( a ) は ( c ) のように定義されている。この ROC 曲線の AUC(Area Under the Curve)は約 ( d ) である。



- |   |        |                  |                  |          |
|---|--------|------------------|------------------|----------|
| 1 | (a)感度  | (b) $TN/(FP+TN)$ | (c) $TP/(TP+FN)$ | (d)0.762 |
| 2 | (a)再現率 | (b) $TP/(TP+FN)$ | (c) $FP/(TP+TN)$ | (d)0.762 |
| 3 | (a)感度  | (b) $TN/(FP+TN)$ | (c) $TP/(TP+FN)$ | (d)0.238 |
| 4 | (a)再現率 | (b) $TP/(TP+FN)$ | (c) $FP/(TP+TN)$ | (d)0.238 |

問 37

確率変数  $X$  の確率密度関数  $f(x)$  が左図のように、累積密度関数  $F(x)$  が右図のように与えられているとする。また、この確率変数  $X$  の 1 次モーメントは  $E[X]=11.20$ 、2 次モーメントは  $E[X^2]=182.40$  であるとする。この確率変数の最頻値、中央値、平均値の大小関係として、もっとも適切なものを選択肢の中から一つ選べ。



- 1 最頻値 < 中央値 < 平均値
- 2 中央値 < 平均値 < 最頻値
- 3 最頻値 < 平均値 < 中央値
- 4 中央値 < 最頻値 < 平均値

問 38

以下の空欄( a )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。

母平均 0、母分散 1 の母集団から得た  $n$  個の互いに独立な標本観測値の標本平均を考える。中心極限定理によると、この標本平均の分布は、平均 0、標準偏差( a )の正規分布に漸近する。

- 1 1
- 2  $1/\sqrt{n}$
- 3  $1/n$
- 4  $1/n^2$

問 39

人工ニューラルネットワークに関する以下の説明文において、( a )、( b )、( c )に入る語および数式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。

分類問題に適用可能な人工ニューラルネットワークは入力層、中間層、出力層からなる。中間層は隠れ層とも呼ばれる。各層を構成するニューロンにおいて、入力信号の総和を出力信号に変換する関数を( a )といい、入力層や中間層ではシグモイド関数、ReLU(rectified linear unit)などの( b )を用いる。また、多クラス分類を考える場合、出力層ではソフトマックス関数( c )を用いることが多い。ここで、 $z_i$ は  $n$  個のニューロンがある出力層の  $i$  番目のニューロンへの総入力とし、 $y_i$ はそのニューロンの出力を表すものとする。また、 $\exp$  は指数関数を表すものとする。

- 1 (a)不活性化関数      (b)線形関数      (c)  $y_i = \frac{1}{1+\exp(-z_i)}$
- 2 (a)活性化関数      (b)線形関数      (c)  $y_i = \frac{\exp(z_i)}{\sum_{j=1}^n \exp(z_j)}$
- 3 (a)不活性化関数      (b)非線形関数      (c)  $y_i = \frac{1}{1+\exp(-z_i)}$
- 4 (a)活性化関数      (b)非線形関数      (c)  $y_i = \frac{\exp(z_i)}{\sum_{j=1}^n \exp(z_j)}$

問 40

$k$  平均法に関する次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1  $k$  平均法は、各クラスタ内のデータの分散の和が小さくなるような、データの  $k$  個のクラスタへの分割を求めるヒューリスティック手法である。
- 2  $n$  個のデータを  $k$  平均法によってクラスタリングする計算量は  $O(kn^2)$  である。
- 3  $k$  平均法では、クラスタ数  $k$  を事前に与える必要がある。
- 4  $k$  平均法の結果は、初期クラスタの選ばれ方によって異なる場合がある。

問 41

次世代シーケンサのマッピング、すなわち、次世代シーケンサにより得られたリードを参照ゲノム配列にマッピングする際、1つのリード（配列）を分割し、分割した部分配列をそれぞれ異なる位置にマッピングする方法がある。これは一般にスプリットアラインメントと呼ばれる。スプリットアラインメントを用いるのに適している場合として、もっとも不適切なものはどれか。選択肢の中から一つ選べ。

- 1 ゲノムシーケンシングにおいて、ゲノムに1塩基置換がある。
- 2 ゲノムシーケンシングにおいて、ゲノムに大規模な欠失がある。
- 3 RNA-seq解析において、RNA スプライシングが起きている。
- 4 RNA-seq解析において、融合遺伝子が存在する。

問 42

近年、参照ゲノム配列や参照トランスクリプトーム配列へのマッピングを行わずに、RNA シーケンシングデータから遺伝子発現量を定量する方法が提案されている。この方法では、各遺伝子配列に含まれる長さ  $k$  の短い部分配列( $k$ -mer)とリードに含まれる  $k$ -mer の組成を比較することで、高速に遺伝子発現量を定量できるが、その際どの程度正しく配列を対応づけられるかが問題となる。これについて述べた以下の文章において、( a ) ~ ( c ) 内に入る語句の組合せとしてもっとも適切なものはどれか。選択肢の中から一つ選べ。

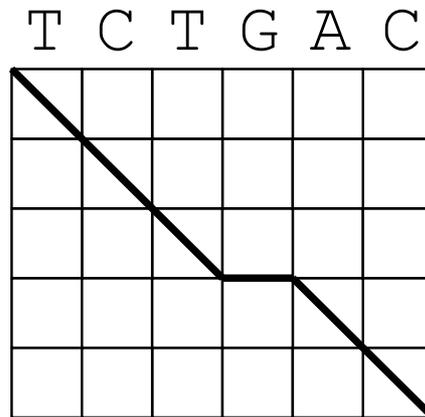
$k$ -mer は塩基配列では ( a ) 種類ある。 $k$ -mer を短くすると検出感度（正しい配列をもれなく検出する割合）は ( b ) と予想される。一方、 $k$ -mer を長くすると偽陽性のヒットが ( c ) と予想される。

- |   |           |          |         |
|---|-----------|----------|---------|
| 1 | (a) $4^k$ | (b) 高くなる | (c) 減る  |
| 2 | (a) $k^4$ | (b) 低くなる | (c) 減る  |
| 3 | (a) $4^k$ | (b) 高くなる | (c) 増える |
| 4 | (a) $k^4$ | (b) 低くなる | (c) 増える |

問 43

ある配列を、配列 TCTGAC とペアワイズアラインメントしたところ、最適パスが下図の太線となるような結果が得られた。ただし、ここではプリン塩基間(A、G)、およびピリミジン塩基間(C、T)の置換よりは、プリン塩基とピリミジン塩基間の置換の方により大きなペナルティが与えられており、さらにギャップにはそれよりも大きなペナルティが与えられているものとする。このとき、アラインメントされた配列として不適切なもの（図のアラインメントパスより高いスコアをとるアラインメントが存在する配列）を選択肢の中から一つ選べ。ただし、選択肢の配列の 1, 4, 5 文字目は、いずれも対象配列の 1, 5, 6 文字目と一致しているため、これらは固定して、アラインメントの可能性としては以下の 3 つを考えればよい。

TCTGAC    TCTGAC    TCTGAC  
 T-XXAC    TX-XAC    TXX-AC



- 1 TGTAC
- 2 TCCAC
- 3 TCAAC
- 4 TACAC

問 44

次世代シーケンサによって得られたデータを、リファレンス配列上にマッピングするためのソフトウェアとして、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 Bowtie
- 2 BWA
- 3 SOAP (SOAP-aligner)
- 4 MAFFT

問 45

以下の ( a ) ~ ( d ) 内に入る数値・語句の組み合わせとしてもっとも適切なものはどれか。選択肢の中から一つ選べ。

ヒトゲノム 2 本鎖における各塩基のおおよその出現頻度は、A:0.3, C: 0.2, G:( a ), T:( b )である。この情報をもとに、2 連続塩基の出現頻度を考えると、CG の出現頻度の期待値は( c )である。この値は、実際のヒトゲノム 2 本鎖における CG の出現頻度よりも ( d )。

- 1 (a)0.2 (b)0.3 (c)0.04 (d)小さい
- 2 (a)0.2 (b)0.3 (c)0.04 (d)大きい
- 3 (a)0.3 (b)0.2 (c)0.06 (d)小さい
- 4 (a)0.3 (b)0.2 (c)0.06 (d)大きい

問 46

以下は大腸菌のゲノム配列の一部である。この中から開始コドンが ATG であり、終止コドンが TGA である ORF を抽出し、そのアミノ酸配列の長さを選択肢から一つ選べ。

5' -TGCCCGTTTCAGGAAGTGCGCCACCAACCTTTCAGTACGAAAATTGCTTTCATTGTCGAT-3'  
3' -ACGGGCAAAGTCCTTCACGCGGTGGTTGGAAAGTCATGCTTTTAACGAAAGTAACAGCTA-5'

- 1 12 aa
- 2 14 aa
- 3 15 aa
- 4 45 aa

問 47

新規のゲノム配列を決定する際に、あらかじめ対象となる生物の大体のゲノムサイズを知っておきたいことが多い。そうした際にゲノムサイズを調査もしくは推定する方法として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 フローサイトメトリー法により DNA 含量を測定する。
- 2 次世代シーケンサで得られたゲノムの断片配列集合における  $k$ -mer の頻度分布を調べる。
- 3 次世代シーケンサで得られたゲノムの断片配列集合における GC 含量と GC skew を計算する。
- 4 DDBJ、ENA、GenBank、C-value database や Animal Genome Size Database などのデータベースに対して、目的の生物種もしくは十分に近縁な生物種が登録されていないか検索する。

#### 問 48

相同性検索によって新規に配列決定された遺伝子の機能を推定する際、配列類似性をもっとも高いと評価された遺伝子のアノテーションに基づいて推定することが多いが、それだけでは正しく推定できない可能性がある。以下に挙げる操作のうち、特に偽陽性に基づく誤った推定を避ける工夫として最も不適切なものを一つ選べ。

- 1  $E$  値( $E$ -value)の閾値を低めに設定して、それより  $E$  値が高い検索結果は捨てる。
- 2 アラインメントを確認して、アラインされた領域が短すぎる検索結果は捨てる。
- 3 上位のヒットが機能未知である場合、下位のヒットのアノテーションを参考にする。
- 4 クエリ配列の中で、アミノ酸組成が偏った領域（低複雑度領域）をマスクして検索する。

#### 問 49

局所的アラインメントにおいては、用いるスコア体系によって、得られるアラインメントの長さやアラインメント中で一致する文字の割合（一致度）が変わってくる。いま、アラインメントのスコアを、一致  $+a$ 、不一致  $-b$ 、ギャップ  $-g$  ( $a, b, g$  は正の整数)で定義し、ギャップは  $g > b$  となるような値で固定するとする。局所的アラインメントとして、なるべく一致度が高くなるような短い領域を得たい場合は、 $a, b$  をそれぞれどのように設定するのがよいか。もっとも適切なものを選択肢の中から一つ選べ。

- 1  $a$  に対して  $b$  を相対的により小さくする。
- 2  $a$  に対して  $b$  を相対的により大きくする。
- 3  $a$  と  $b$  を同じ割合で大きくする。
- 4  $a$  と  $b$  を同じ割合で小さくする。

問 50

以下の人名のついた手法は、いずれもバイオインフォマティクスにおける古典的な手法である。それぞれの用途を適切に対応づけたものを選択肢の中から一つ選べ。

NW ニードルマン・ブンシュ(Needleman-Wunsch)法

SW スミス・ウォーターマン(Smith-Waterman)法

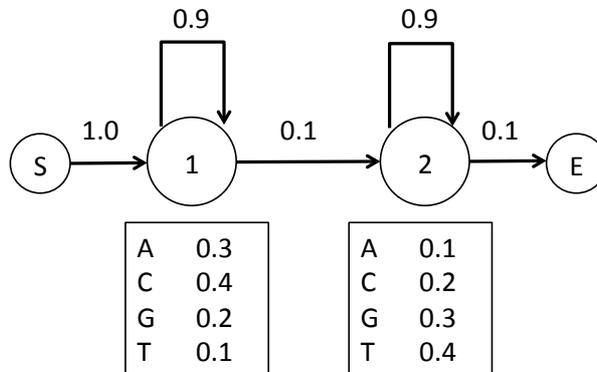
KD カイト・ドゥーリトル (Kyte-Doolittle) 法

CF チョウ・ファスマン(Chou-Fasman)法

選択肢	タンパク質 二次構造予測	疎水性プロット (膜貫通部位予測)	大域的配列アライン メント	局所的配列アライン メント
1	CF	KD	NW	SW
2	KD	CF	NW	SW
3	CF	KD	SW	NW
4	KD	CF	SW	NW

(問 51・問 52)

以下の図は、固有の塩基出現確率を持つ 2 つの状態 1, 2 を持つ、DNA 配列を記述する隠れマルコフモデルを表している。ただし、S と E はそれぞれ開始、終了状態を表す。これについて、以下の問いに答えよ。



問 51

配列 AC がこのモデルから出力される (状態 1 で A、状態 2 で C を出力して終了する) 確率はいくらか。選択肢の中から適切なものを一つ選べ。

- 1 0.06
- 2 0.0081
- 3 0.00108
- 4 0.0006

問 52

このモデルに配列 ACAGCTGT を当てはめたとき、もっとも尤度が高い状態遷移のパスはどれか。選択肢の中から一つ選べ。ただし、以下の表を参考にしてよい。

(参考表)

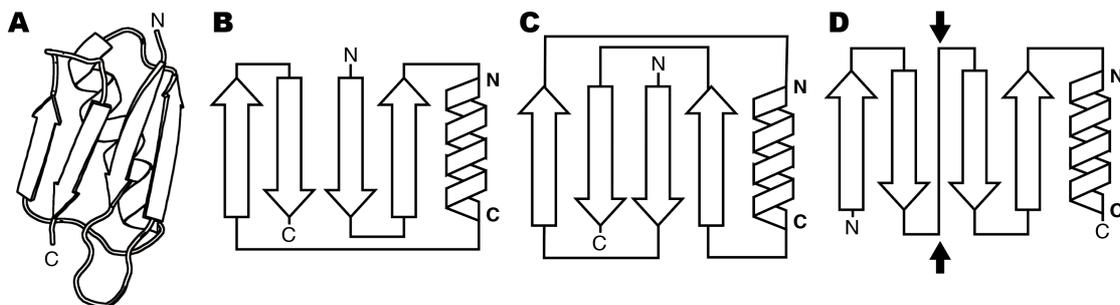
状態	A	C	A	G	C	T	G	T
1	0.3	0.4	0.3	0.2	0.4	0.1	0.2	0.1
2	0.1	0.2	0.1	0.3	0.2	0.4	0.3	0.4

【選択肢】

- A C A G C T G T -
- 1 S 1 1 1 2 2 2 2 2 E
  - 2 S 1 1 1 1 2 2 2 2 E
  - 3 S 1 1 1 1 1 2 2 2 E
  - 4 S 1 1 1 1 1 1 2 2 E

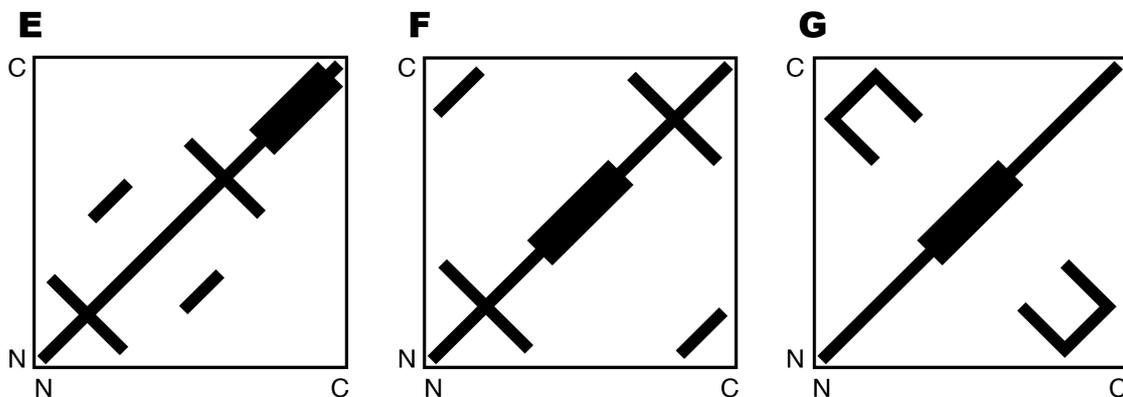
(問 53・問 54)

図 A はあるタンパク質(PDB ID: 2GB1)の立体構造を示している。この構造のトポロジーを表現したものが図 B である。このタンパク質の二次構造要素の向きや位置など三次元的な配置をまったく変えずに、ループをつなぎ替えることで新しいタンパク質をデザインしたい。その例が図 C および D である。これについて問 53 および問 54 に答えよ。



問 53

下図 E、F、G について、トポロジー図 B、C、D に対するコンタクトマップとして適切なものはどれか。適切な組み合わせを選択肢の中から一つ選べ。



- 1 B と F、C と G、D と E
- 2 B と G、C と F、D と E
- 3 B と E、C と G、D と F
- 4 B と F、C と E、D と G

問 54

二次構造要素の向きや位置など三次元的な配置をまったく変えずにループをつなぎ替えるとして、何種類の異なる構造を作れるだろうか。以下の 2 つの場合について、正しい組み合わせを選択肢の中から一つ選べ。

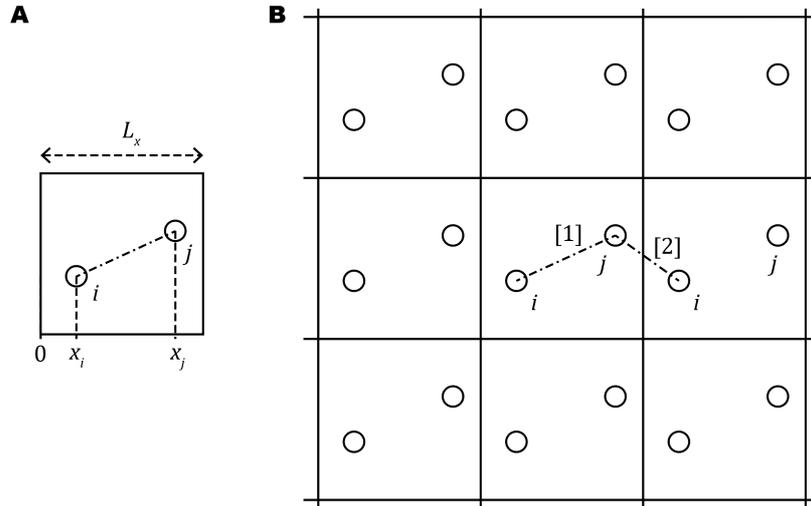
(a) 全ての可能なつなぎ方の数

(b) 図 D ではひとつのループ（黒塗り矢印で示したもの）が、図中の下から上へとつながっている。このような「ループが上下を横切る構造」を禁止した場合の数

- |   |         |        |
|---|---------|--------|
| 1 | (a) 240 | (b) 24 |
| 2 | (a) 120 | (b) 12 |
| 3 | (a) 240 | (b) 12 |
| 4 | (a) 120 | (b) 24 |

問 55

分子シミュレーションでは周期的境界条件がよく用いられる。この条件では全く同じ構造のシミュレーションセルが周期的に並んでいるものと考えられる。つまり図 A の系は、計算上では図 B のように扱われる。このとき原子  $i$  と原子  $j$  の間の距離は、最も近い  $i - j$  ペアの距離が採用される。すなわち、下図 B においては [1] でなく [2] が原子  $i$  と原子  $j$  の間の距離となる。



この距離を求める処理を実装したい。まずは  $x$  軸のみについて考えるものとして、原子  $i$  と原子  $j$  のもっとも近いペアについて  $x$  軸上の距離  $r_x$  を求める。下記のプログラム（擬似コード）において、空欄（ a ）にあてはまる処理として正しいものを選択肢の中から一つ選べ。なお、原子  $i$  と原子  $j$  の  $x$  座標をそれぞれ  $x_i, x_j$  とし、セルの  $x$  軸方向の長さを  $L_x$  とする。

【擬似コード】

```

 $r_x = \text{abs}(x_j - x_i);$ 
 $\text{if}(r_x > ( a )) \quad r_x = L_x - r_x;$ 

```

【選択肢】

- 1  $0.5 * L_x$
- 2  $L_x$
- 3  $2 * L_x$
- 4  $L_x * L_x$

問 56

下図はある転写因子のアミノ酸配列を示している。三角はリン酸化部位を示している。部位 ( a ) ~ ( c ) の特徴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。

```

          1234567890 1234567890 1234567890
271  QSSFNSLQRV (a)▼▼ PSYDSFDSED YPAALPNHKP
301  KGTFKDYVRD RADLNKDKPV IPAAALAGYT
331  GSGPIQLWQF (b) LLELLTDKSC QSFISWTGDG
361  WEFKLSDPDE VARRWGKRKN KPKMNYEKLS
391  (c) RGLRYYYDKN I IHKTAGKRY VYRFVCDLQS
421  LLGYTPEELH AMLDVKPDAD E

```

- |   |                    |                    |                    |
|---|--------------------|--------------------|--------------------|
| 1 | (a) $\alpha$ ヘリックス | (b) DNA 結合部位       | (c) 天然変性領域         |
| 2 | (a) 天然変性領域         | (b) $\alpha$ ヘリックス | (c) DNA 結合部位       |
| 3 | (a) DNA 結合部位       | (b) $\alpha$ ヘリックス | (c) 天然変性領域         |
| 4 | (a) 天然変性領域         | (b) DNA 結合部位       | (c) $\alpha$ ヘリックス |

問 57

タンパク質の立体構造を記述するファイルフォーマットに関する以下の記述についてもっとも不適切なものを選択肢の中から一つ選べ。

- 1 PDB フォーマットでは、行内の各要素は空白で区切られて記述される。
- 2 mmCIF フォーマットでは、データは階層性のあるタグで分類される。
- 3 PDBML フォーマットでは、XML と同様にタグを用いてデータを記述する。
- 4 現在では mmCIF 形式が標準フォーマットである。

問 58

以下の図はある二つのタンパク質の立体構造を模式的に示したものである。この図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。

A



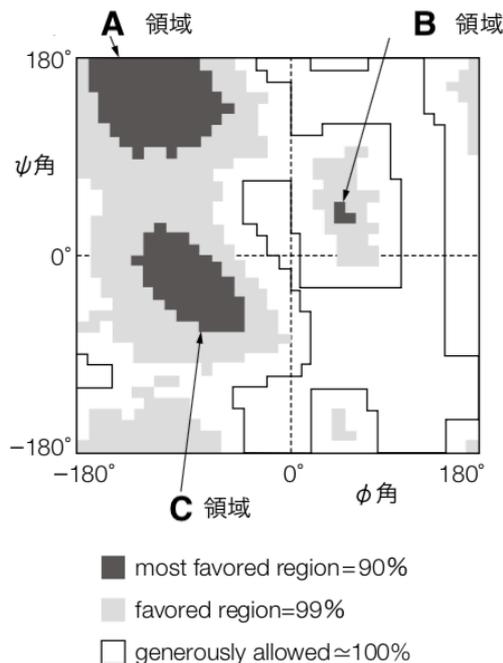
B



- 1 A と B は同じフォールドに属する。
- 2 A は all- $\alpha$  型、B は all- $\beta$  型である。
- 3 A、B ともにシングルドメインタンパク質である。
- 4 A は二量体、B は単量体である。

問 59

ラマチャンドランマップは、横軸に  $\phi$  角（主鎖の窒素原子と  $\alpha$  炭素原子の間の二面角）、縦軸に  $\psi$  角（主鎖のカルボニル基の炭素原子と  $\alpha$  炭素原子の間の二面角）をとり、対応する点にそれぞれのアミノ酸残基をプロットする。このマップは、構造が明らかにされているタンパク質内のアミノ酸による統計的分布から、最も好まれる **most favored** 領域（90%以上のアミノ酸はこの領域に入る。図の濃灰色領域）、**favored** 領域（99%以上のアミノ酸はこの領域に入る。図の薄灰色領域）、**generously allowed** 領域（図の実線内）に分かれる。**most favored** 領域は大きく3つに分かれており、それぞれ  $\alpha$  ヘリックス領域、 $\beta$  シート領域、左巻きらせん（ターン）領域に対応する。ラマチャンドランマップ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 現在までに立体構造が明らかにされている全タンパク質についてこのマップを作成すると、このマップの作成法が提案された1963年当時より、**most favored** 領域、**favored** 領域、そして、**generously allowed** 領域は広くなる。
- 2 任意のアミノ酸の種類ごとにラマチャンドランマップを描くと、グリシンやプロリンなどのアミノ酸のマップでは、他のアミノ酸とは異なる特徴的な分布となる。
- 3 **A** は  $\alpha$  ヘリックス領域、**B** は左巻きらせん領域、**C** は  $\beta$  シート領域に対応する。
- 4 タンパク質の立体構造予測では、予測された立体構造についてこのマップを作成し、局所的な構造の予測精度を確認するために用いられている。

#### 問 60

解析対象タンパク質と相同な配列を含むマルチプルアラインメントを構築できると、各アラインメントサイトにおいて進化的な保存度のスコアを算出できる。サイトの保存度は構造の維持や機能の発揮に重要なアミノ酸を推定するための良い指標の一つである。さらに、保存度が高いサイトに対応するアミノ酸（保存残基）の立体構造上での位置を考えることで、より詳細な情報を得られることがある。このことに関連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 立体構造内部に埋もれている保存残基は立体構造形成のために重要度が高い傾向にある。
- 2 立体構造上の分子表面における保存残基の探索は、酵素の活性部位やタンパク質間相互作用インターフェイスを構成するアミノ酸の予測に利用されている。
- 3 基質の認識に重要な働きをするアミノ酸は、基質が接触可能な分子表面に常に位置する必要があり、立体構造内部に埋もれることはない。したがって、酵素の立体構造表面に位置する保存残基を捉えることで基質認識部位を予測することができる。
- 4 立体構造上の分子表面における、保存残基の集積度の高さを統計的に評価することで、機能部位としてはたらく領域を予測する手法がある。

#### 問 61

次世代シーケンス解析で扱うファイルタイプ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 **BAM**: ゲノムに次世代シーケンサなどから出力された配列をマッピングした時に得られる配列のアラインメント情報に関するファイルタイプ。
- 2 **VCF**: エクソーム解析などで変異情報を記述するときのファイルタイプ。
- 3 **GTF**: ChIP-seq などのピークを検出した時に出力されるファイルタイプ。
- 4 **FASTQ**: DNA などの塩基配列とそのクオリティスコアを1つのファイルに保存するためのファイルタイプで、クオリティスコアは1文字のASCIIコードで表されている。

問 62

以下のメンデル遺伝についての記述の ( a ) ~ ( c ) に入る語句・記号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。

メンデルの研究したエンドウマメの種子の形は丸(R)としわ(r)、色は黄(Y)と緑(y)を表現型とする対立遺伝子によって支配されている(括弧内は対立遺伝子の記号とする)。純系の丸・緑と純系のしわ・黄の個体を掛け合わせた F1 の遺伝子型は ( a ) であるが、これらの種子の表現型はすべて丸・黄であったので、しわ・緑が ( b ) 形質であることがわかる。さらに F1 どうしをかけ合わせた F2 の表現型分離比(丸・黄:丸・緑:しわ・黄:しわ・緑)が 9:3:3:1 に近い値になったことから、これらの遺伝子座はメンデルの ( c ) の法則に従うことがわかる。

選択肢	a	b	c
1	RRYY	潜性 (劣性)	独立
2	RRYY	顕性 (優性)	連鎖
3	RrYy	潜性 (劣性)	独立
4	RrYy	顕性 (優性)	連鎖

問 63

ある 2 倍体生物の遺伝子座 A には、A<sub>1</sub> と A<sub>2</sub> の 2 種類の対立遺伝子が存在している。この生物の集団について遺伝子型を調査したところ、以下の表に示す結果が得られた。

遺伝子型	A <sub>1</sub> A <sub>1</sub>	A <sub>1</sub> A <sub>2</sub>	A <sub>2</sub> A <sub>2</sub>
個体数	10	80	10

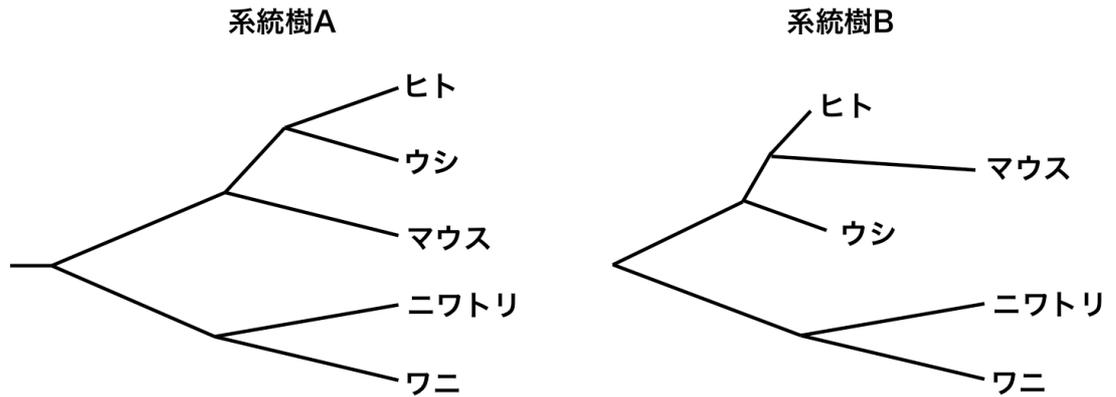
この結果から、対立遺伝子 A<sub>1</sub> と A<sub>2</sub> の頻度、およびこの集団で遺伝子座 A にハーディー・ワインベルク平衡が成立しているか否かについて、もっとも適切な組み合わせを選択肢の中から一つ選べ。ハーディー・ワインベルク平衡の検定は、 $\chi^2$  検定で有意水準を 1% に設定して行うものとする(P 値 = 0.01 の自由度 1 の  $\chi^2$  値は 6.6349 である)。

選択肢	A <sub>1</sub> 頻度	A <sub>2</sub> 頻度	ハーディー・ワインベルク平衡
1	0.5	0.5	成立している
2	0.9	0.1	成立していない
3	0.9	0.1	成立している
4	0.5	0.5	成立していない

(問 64・問 65・問 66)

問 64

5種の生物（ヒト、ウシ、マウス、ニワトリ、ワニ）由来の相同遺伝子について進化距離（座位あたりの置換数）を求め、異なる方法を使用して図に示す2通りの系統樹AとBを作成した。系統樹AおよびBの作成法、およびそのように判断する理由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。



選択肢	系統樹 A 作成法	系統樹 B 作成法	判断する理由
1	UPGMA 法	NJ 法	系統樹 A にはルートがあり、系統樹 B にはルートがない
2	UPGMA 法	NJ 法	系統樹 A は進化速度が一定でなく、系統樹 B は一定である
3	NJ 法	UPGMA 法	系統樹 A にはルートがなく、系統樹 B にはルートがある
4	NJ 法	UPGMA 法	系統樹 A は進化速度が一定であり、系統樹 B は一定でない

問 65

問 64 に示した系統樹 A のニューウィック (Newick) 表現として、不適切なものを選択肢の中から一つ選べ。

- 1 (((ヒト, ウシ), マウス), (ニワトリ, ワニ));
- 2 ((ニワトリ, ワニ), ((ヒト, ウシ), マウス));
- 3 ((ヒト, ウシ), マウス, (ニワトリ, ワニ));
- 4 (((ウシ, ヒト), マウス), (ニワトリ, ワニ));

問 66

問 64 に示した 2 つの系統樹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 この遺伝子については、UPGMA 法と NJ 法でトポロジーの異なる系統樹が得られる。
- 2 系統樹作成法には大きく分けて、進化距離を用いる距離行列法と、配列データを直接用いる文字置換利用法があり、UPGMA 法と NJ 法はいずれも文字置換利用法である。
- 3 2 つの方法で求めた系統樹間に差異があるのは、UPGMA 法ではすべての枝で進化速度が一定であることを暗に仮定しているのに対し、NJ 法では進化速度の一定性を仮定しないためである。
- 4 2 つの方法で求めた系統樹間に差異があるのは、この遺伝子の進化速度がマウスの系統で近縁種と比べて上昇していることに起因すると考えられる。

問 67

塩基置換に関する以下の記述について、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 トランスバージョン変異（プリンとピリミジン間の変異）よりも、トランジション変異（プリン間またはピリミジン間の変異）の方が起こりにくい。
- 2 塩基置換を HGVS (Human Genome Variation Society) 表記法で c.411A>T のように表すことがあるが、これは coding DNA の 411 番目の塩基が A から T に変わっていることを示す。
- 3 アラインメントをしたときに同じ塩基配列であっても、同じ塩基で 2 回以上置換されてもとに戻っている場合もありえる。
- 4 遺伝子のサイトあたりの非同義置換数( $dN$ )とサイトあたりの同義置換数( $dS$ )の値は、通常  $dN < dS$  である。

問 68

連鎖解析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 連鎖不平衡とは、複数の遺伝子座の対立遺伝子の特定の組み合わせ頻度が有意に高くなる現象である。
- 2 連鎖不平衡にある遺伝子座は同一染色体に存在するが多いが、交配がランダムでない集団などでは、異なる染色体上に存在する遺伝子座の間でも連鎖不平衡が認められる。
- 3 連鎖不平衡にある複数の遺伝子座の対立遺伝子の特定の組み合わせがディプロタイプであり、相同染色体間でのディプロタイプの組み合わせをハプロタイプとよぶ。
- 4 LOD(logarithm of odds)スコアを用いた家系分析を行うことにより、遺伝的疾患の原因遺伝子が連鎖地図上でマーカー遺伝子からどの程度離れた位置にあるかを推定できる。

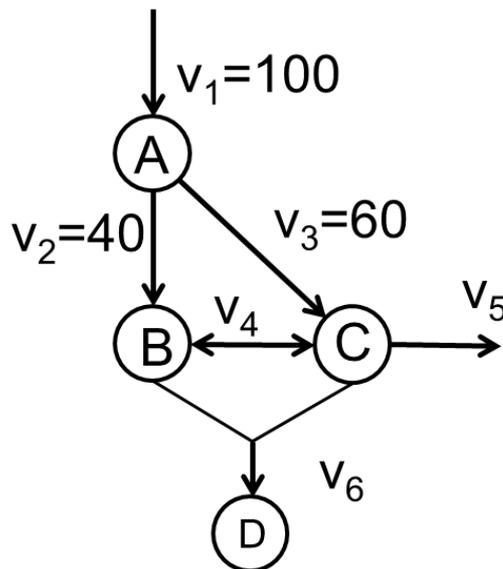
問 69

図は代謝物 (A, B, C, D) 間の反応を示している。定常状態において、反応  $v_1$  から  $v_6$  の反応式と  $v_1$  から  $v_3$  のモル流束値は表のように与えられる。 $v_6$  の定常状態流束のとりうる上限として正しいものを選択肢の中から一つ選べ。 $v_6$  の反応では、1 モルの B と 1 モルの C が反応して、1 モルの D が生成する。

表 代謝反応

反応	モル流束値	反応式	反応
$v_1$	100	$\rightarrow 1 A$	不可逆
$v_2$	40	$1 A \rightarrow 1 B$	不可逆
$v_3$	60	$1 A \rightarrow 1 C$	不可逆
$v_4$		$1 B \leftrightarrow 1 C$	可逆
$v_5$		$1 C \rightarrow$	不可逆
$v_6$		$1 B + 1 C \rightarrow 1 D$	不可逆

図 代謝ネットワークマップ



- 1 40
- 2 50
- 3 60
- 4 100

問 70

遺伝子の転写速度 $v$ が転写因子濃度 $y$ を用いて、

$$v = V \frac{K}{K + y}$$

と表されている。ただし、 $V$ は最大転写速度定数、 $K$ は解離定数である。この式に関する記述として不適切なものを選択肢の中から一つ選べ。

- 1 転写速度を最大にする条件は $y = 0$  である。
- 2 最大転写速度の4分の1の速度を与えるとき、 $y = 4K$ である。
- 3  $y$ が一定のとき、 $K$ の値が大きくなると転写速度 $v$ は速くなる。
- 4  $y$ が増加すると、遺伝子の転写は抑制される。

問 71

生物ネットワークに関する以下の記述についてもっとも不適切なものを選択肢の中から一つ選べ。

- 1 生物ネットワークには代謝ネットワーク、遺伝子制御ネットワーク、タンパク質間相互作用ネットワークなどがある。
- 2 現実に観測されたネットワークにおいて、統計学的に有意に多く観測される部分構造をネットワークモチーフと呼ぶ。
- 3 統計学的にネットワークモチーフを見出す際、帰無仮説に相当するネットワークはスケールフリーネットワークがよく用いられる。
- 4 遺伝子制御ネットワークではフィードフォワードループが特徴的に見出されている。

問 72

細胞濃度  $X$  の時間変化が比例定数  $k (>0)$ を用いて、

$$dX/dt = kX$$

と与えられるとき、細胞濃度が2倍になるために要する時間（ダブリングタイム）の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 ダブリングタイムは細胞濃度  $X$  に依存し、 $X$  が大きいほど短くなる。
- 2 ダブリングタイムは細胞濃度  $X$  に依存し、 $X$  が大きいほど長くなる。
- 3 ダブリングタイムは比例定数  $k$  に依存し、 $k$  が大きいほど長くなる。
- 4 ダブリングタイムは比例定数  $k$  に依存し、 $k$  が大きいほど短くなる。

問 73 全員正解としました

A および B の 2 つの試料を対象に短鎖シーケンサ（リード長 75bp）で RNA-Seq 解析を行った。試料 A において、転写産物 X および転写産物 Y にマッピングされたリードの数はそれぞれ 300 と 500 であった。試料 B において、転写産物 X および転写産物 Y にマッピングされたリードの数はそれぞれ 900 と 1,800 であった。また、試料 A および試料 B において取得したリードの総数はそれぞれ 10,000,000 と 30,000,000 であった。転写産物 X および転写産物 Y の長さはそれぞれ 1,000 と 2,000 である。転写産物の発現に関する推定として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 転写産物 X の発現量は、試料 B よりも試料 A の方が大きい。
- 2 試料 A において、転写産物 X は転写産物 Y よりも多く発現している。
- 3 転写産物 Y の発現量は、試料 A よりも試料 B の方が大きい。
- 4 試料 B において、転写産物 Y は転写産物 X よりも多く発現している。

\*本問は問題文の推敲課程で、手計算で計算し易い数字に修正しましたが、その結果選択肢 1 と 4 において、ノーマライズ後の発現量に違いがなくなるために、正解選択肢が 2 つある不適切な問題となっていました。そのためこの問題は全員正解とします。

問 74

次に示したデータベース名とその内容の組み合わせの中で、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 GenBank: 塩基配列
- 2 UniProt: アミノ酸配列
- 3 JASPAR: タンパク質間相互作用
- 4 Gene Expression Omnibus (GEO): 遺伝子発現

問 75

次世代シーケンサを用いた解析手法の説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 Hi-C 解析は、ゲノム DNA の三次元構造を明らかにすることができる。
- 2 RamDA-Seq 解析は、試料中に存在する生物群の全ゲノム DNA 配列を網羅的に解析し、新規遺伝子資源の探索に応用することができる。
- 3 ChIP-Seq 解析は、特定のタンパク質が結合しているゲノム DNA 領域を同定することができる。
- 4 RNA-Seq 解析は、全転写産物の発現量を網羅的に推定することができる。

問 76

近年、腸内や土壌に含まれる細菌叢の構成は、それぞれヒトの健康状態や農業効率化に大きな影響を及ぼすことが知られている。これら細菌叢の構成を解析する細菌叢解析の一つの手法である 16S Amplicon-Seq 解析について説明した次の文について、( a )～( c )内に入る語句として正しい組み合わせを選択肢の中から一つ選べ。

細菌が持つ 16S rDNA の定常領域を対象としたプライマーを用いて PCR を行い、次世代シーケンサを用いて増幅 DNA に含まれる 16S rDNA 可変領域の塩基配列を決定する。続いて、塩基配列の類似度が高い増幅 DNA 群を( a )として統合し、BLAST などの配列相同性検索プログラムを用いて( b )などの完全長 16S rDNA データベースを照会することで個々の( a )に菌種を帰属し、解析試料の菌種組成を明らかにする。また、UniFrac などの多様性指標を用いて試料間の細菌叢組成類似度を評価するなど、種々の生態学的解析を実施する。これら一連のデータ処理を効率的に実施するための解析パイプラインとして( c )などが報告されている。

- |   |                                      |            |           |
|---|--------------------------------------|------------|-----------|
| 1 | (a) OTU (Operational Taxonomic Unit) | (b) SILVA  | (c) QIIME |
| 2 | (a) ORF (Open Reading Frame)         | (b) SILVA  | (c) GATK  |
| 3 | (a) OTU (Operational Taxonomic Unit) | (b) STRING | (c) GATK  |
| 4 | (a) ORF (Open Reading Frame)         | (b) STRING | (c) QIIME |

問 77

血液中や尿などの生体サンプル内の代謝物を網羅的に測定するメタボローム解析について、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 核磁気共鳴 (NMR) 法によるメタボローム解析は、測定のスループットが高く、多くのサンプルを測定するのに適している。
- 2 液体クロマトグラフィー質量分析 (LC/MS) 法によるメタボローム解析は、極性が比較的高い成分の分析に有効である。
- 3 ガスクロマトグラム質量分析 (GC/MS) 法によるメタボローム解析は、ガス状または気化する成分の分析に有効である。
- 4 NMR 法は LC/MS 法に比べて感度が高く、より少ないサンプルで濃度の低い化合物を測定することができる。

問 78

一細胞 RNA-Seq 解析により  $N$  個の細胞の発現量解析を行い、細胞を発現量により  $k$  個のクラスタに分割し、クラスタ間で発現量の異なる遺伝子を検討した。この際のデータ解析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 クラスタ解析の結果からクラスタ毎に発現量の異なる遺伝子の探索を行うために、すべてのクラスタペアで遺伝子の発現量の  $t$  検定を行った。
- 2 クラスタ解析として、階層的クラスタ解析と  $k$  平均法クラスタ解析を実施して結果を比較し、結果に大きな差は無かったので  $k$  平均法クラスタ解析を採用した。
- 3  $k$  平均法クラスタ解析を行う際に AIC により適切なクラスタ数を決めた。
- 4 クラスタ解析を行う際には、遺伝子の発現量の類似性として、ピアソン相関係数を用いた。

問 79

遺伝子発現量解析を行うために、ある培養細胞から得られる RNA サンプル A および同種の培養細胞に酸化ストレスを与えて得られるサンプル B を RNA-Seq 解析したところ、遺伝子 X に関してはサンプル B 由来の発現量が、サンプル A 由来の発現量に比べて高い値を示した。この結果を確認するため、リアルタイム PCR 法で遺伝子 X の発現量を調べたところ、サンプル B の Ct 値（増幅曲線がある一定のシグナル値を超えるサイクル数）がサンプル A の Ct 値より 2 だけ低かった。これから導かれる結論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 遺伝子 X はサンプル A においてサンプル B の 2 倍量発現している。
- 2 遺伝子 X はサンプル A においてサンプル B の 4 倍量発現している。
- 3 遺伝子 X はサンプル B においてサンプル A の 2 倍量発現している。
- 4 遺伝子 X はサンプル B においてサンプル A の 4 倍量発現している。

問 80

転写制御解析やエピゲノム解析で行われる解析手法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一つ選べ。

- 1 ゲノム上の転写因子結合部位を推定するため、抗転写因子抗体を用いた ChIP-Seq 解析が行われている。
- 2 DNA 上のメチル化部位を推定するために、抗メチル化 DNA 抗体を用いた MeDIP-Seq 解析が用いられる。
- 3 ヒストンの修飾に応じた抗体を使って ChIP-Seq を行うことで、ゲノム上のヒストン修飾部位が推定できる。
- 4 ヌクレオソームが DNA 上にどのように配置されているかを調べるために、バイサルファイトシーケンス法が用いられる。

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

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。

Copyright © 2018 Japanese Society for Bioinformatics. All Rights Reserved.