

**平成29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。**

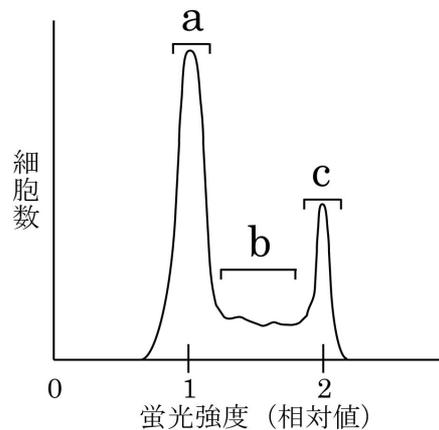
問 1

電気泳動を用いた DNA 分子の分離方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。

- 1 サンプル DNA をあらかじめ制限酵素で断片化しておくことで、特異的なバンドのパターンが得られる。
- 2 DNA は負電荷を帯びており、陽極に向かって移動する。
- 3 ゲル内の DNA 分子の泳動距離は、DNA 分子の長さに比例し、分子の長さが長いものほど泳動時間に応じて長く進む。
- 4 電気泳動された DNA はゲルから回収することも可能である。

(問 2・問 3)

ある細胞集団の核を蛍光色素で染色し、フローサイトメーターを用いて個々の細胞の蛍光強度を測定した。縦軸に細胞数、横軸に蛍光強度をとったグラフを作成したところ、以下の図のようになった。この実験について、以下の問 2 および問 3 に答えよ。



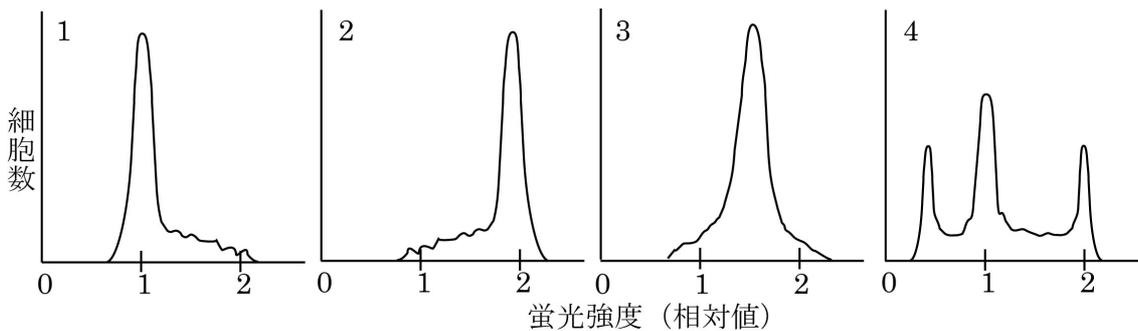
問 2

G<sub>1</sub>期、S期、G<sub>2</sub>期、M期にある細胞は、主に a~c のどの範囲に含まれるか。もっとも適切なものを選択肢の中から一つ選べ。

- |   |                      |        |                      |        |
|---|----------------------|--------|----------------------|--------|
| 1 | G <sub>1</sub> 期 : a | S期 : c | G <sub>2</sub> 期 : b | M期 : b |
| 2 | G <sub>1</sub> 期 : a | S期 : b | G <sub>2</sub> 期 : c | M期 : c |
| 3 | G <sub>1</sub> 期 : b | S期 : c | G <sub>2</sub> 期 : a | M期 : a |
| 4 | G <sub>1</sub> 期 : b | S期 : b | G <sub>2</sub> 期 : c | M期 : a |

### 問 3

有糸分裂の際に形成される紡錘体は、チューブリンタンパク質を取り込んで重合することで、微小管を伸ばしていくことが知られている。問 2 で観察した細胞集団にコルヒチン処理を行うと、蛍光強度と細胞数はどのような分布になるか。適切なものを以下の選択肢の中から一つ選べ。なお、コルヒチンはチューブリン重合阻害剤である。



### 問 4

メンデルは「遺伝子の型の相違が遺伝性の形質の相違を引き起こす」ことに気づいており、このような遺伝子の型の相違をもつ遺伝子を、現在では対立遺伝子（アレル）と呼ぶ。メンデルの法則について書かれた次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

（注：最近になって、遺伝学分野で長らく使われて来た「優性(dominance)」と「劣性(recessive)」を、それぞれ「顕性」と「潜性」と言い換えるのがより良いとの主張が、国内学会からなされている。ただしまだ広く認知されていないことから、この問題では国内高等学校の教科書等でもこれまで採用されてきた「優性」及び「劣性」という言葉を用いた。）

- 1 「配偶子形成過程で個々の対立遺伝子の対は、他の対立遺伝子の対とは独立に配偶子に分配されていく」ことが知られており、独立の法則(law of independent assortment)と呼ばれている。
- 2 「特定の形質に対して、優性の対立遺伝子は劣勢の対立遺伝子よりも集団内での存在比率が高い」ことが知られており、優性の法則(law of dominance)と呼ばれている。
- 3 「遺伝性の形質を示す二つの対立遺伝子は配偶子の形成過程で分離し、別々の配偶子に配分される」ことが知られており、分離の法則(law of segregation)と呼ばれる。
- 4 実際の遺伝様式は、単純なメンデル遺伝学による予想よりも複雑な例が多く知られている。

### 問 5

進化していない集団の遺伝子プールは、ハーディ・ワインベルグの法則により記述することが可能である。ハーディ・ワインベルグの法則が成り立つ仮想集団が満たすべき条件のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 ハーディ・ワインベルグの法則が成り立つ仮想集団では、突然変異は起きないこととする。
- 2 ハーディ・ワインベルグの法則が成り立つ仮想集団では、ランダムな交配は起きないこととする。
- 3 ハーディ・ワインベルグの法則が成り立つ仮想集団では、自然選択は起きないこととする。
- 4 ハーディ・ワインベルグの法則が成り立つ仮想集団では、非常に大きな集団サイズを仮定している。

### 問 6

エピジェネティック遺伝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 ヒストンがアセチル化されると、クロマチン構造が弛緩し、転写関連タンパク質が結合しやすくなることが知られている。
- 2 ヒストンがメチル化されると、クロマチンの凝縮が促進されることが知られている。
- 3 DNA のメチル化は、動物と真菌でのみ観察されており、植物では生じないことが知られている。
- 4 一卵性双生児であっても、エピジェネティックな修飾には違いがあることが知られている。

### 問 7

動物の内分泌系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 松果体と甲状腺は、いずれもヒトの主要な内分泌腺である。
- 2 ステロイドホルモンは親水性であり、細胞膜を容易に通過できる。
- 3 ヒトの性ホルモンは、主に雄の精巣と雌の卵巣から分泌されるが、副腎からも少量の性ホルモンを分泌する。
- 4 脳下垂体前葉は内分泌腺の一種であり、視床下部からのシグナルに反応してホルモンを生産し分泌する。

#### 問 8

動物の神経系について述べた次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 クラゲやイソギンチャクといった刺胞動物にも神経細胞が存在するが、カイメン動物には神経細胞が存在しない。
- 2 神経組織において、神経インパルスを伝達する細胞をニューロン（神経細胞）、支持細胞としてはたらく細胞をグリアと呼ぶ。
- 3 ニューロンが次のニューロンにシグナルを伝達する部位をシナプスと呼び、シナプスにおいて神経伝達物質が次の細胞に情報を伝える。
- 4 脊椎動物の軸索は複数のニューロンが一本に束ねられたもので、その周りがミエリン鞘と呼ばれる電氣的に絶縁された層に覆われている。

#### 問 9

細菌の DNA 複製に関して書かれた次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 DNA の複製は、複製起点と呼ばれる特異的な塩基配列を持つ短い DNA 領域から開始される。環状の DNA をもつ細菌において、複製起点は対置するように二箇所存在する。
- 2 DNA 複製を開始するタンパク質は、複製起点の塩基配列を認識して結合し、二本鎖を分離する。分離された領域は「複製バブル」と呼ばれる。
- 3 二本鎖の分離が進んでいる両端は「複製フォーク」と呼ばれる。複製フォークの前方においては、二重らせん DNA 領域のねじれがきつくなって緊張が高まるが、トポイソメラーゼと呼ばれる酵素が DNA 鎖を一時切断し、回転させてから再結合することで緊張を緩和させている。
- 4 DNA ポリメラーゼは 5' から 3' へ方向にしか伸長することができない。逆側の方向への複製、すなわちラギング鎖の複製にはプライマーゼと DNA ポリメラーゼ III が協調して「岡崎フラグメント」を形成することで行われる。

#### 問 10

真核生物におけるメッセンジャーRNA(mRNA)のスプライシングについて書かれた次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 mRNA 前駆体からのイントロンの切り出しは、スプライソソームと呼ばれる大きな構造体が担う。スプライソソームは数個の核内低分子リボ核タンパク質(snRNP)と他のいくつかのタンパク質によって構成されている。
- 2 イントロンは切り出され方の種類や自己触媒能などによって、いくつかのタイプに分類されている。スプライソソームによって担われるイントロンは、スプライソソーム型イントロンと呼ばれる。
- 3 スプライソソーム型イントロンは、その多くが GT-AG 則と呼ばれる特定の配列をもつ。すなわちその 5' 末端に GT が、3' 末端に AG が配置されている。
- 4 mRNA 前駆体はそのまま核外に輸送され、エキソンの切り出しは、細胞質で行われる。

#### 問 11

細胞内におけるタンパク質の発現機構について書かれた次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 翻訳とはメッセンジャーRNA(mRNA)の情報を用いて、ポリペプチドを合成することである。
- 2 ポリペプチドとは、多数の核酸がペプチド結合によってつながった重合体のことである。
- 3 翻訳を担う細胞内小器官はリボソームと呼ばれ、リボソーム RNA(rRNA)と呼ばれる RNA と、タンパク質の複合体で構成されている。
- 4 ポリペプチド鎖合成に必要なエネルギーは、GTP の加水分解によって供給される。

#### 問 12

生体内や生体間には細胞同士が情報をやり取りするために様々なシグナル伝達の経路が存在する。シグナル伝達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 動物の発生において分泌性のシグナル分子が、細胞分化に関わることが知られている。その一つである WNT タンパク質は、ほとんどの動物のゲノムにコードされ、背腹の軸形成など、発生における様々な役割を担っている。
- 2 細胞性粘菌は化学シグナルを使って栄養物があるかどうかを互いに情報連絡し合っている。食物が欠乏すると、飢えた細胞が、近くの細胞に凝集するように刺激する分子を分泌する。
- 3 動物や植物は、ホルモンと呼ばれる化学物質を長距離のシグナル伝達に使っている。動物では呼吸器系を通してホルモンが伝達するのに対し、植物ホルモンは導管や細胞を通るほか、気体として空気中を通して標的に到達することもある。
- 4 神経系を伝わるシグナル伝達では、電気シグナルが化学シグナルに変換される。

問 13

ゲノム編集技術に関する次の記述を読み、もっとも不適切なものを選択肢の中から一つ選べ。

ゲノム編集技術と呼ばれる技術として、TALEN（転写活性化様エフェクターヌクレアーゼ）、ZFNs（ジンクフィンガーヌクレアーゼ）、CRISPR/Cas9 の三つがよく知られている。いずれの技術においても、標的 DNA の近傍に結合するための DNA 結合ドメインと、標的 DNA を切るなどの操作をする機能ドメインをもつ。

- 1 ZFNs においては、DNA 結合ドメインは 3 塩基を認識するジンクフィンガーモチーフ数個で構成されている。
- 2 TALEN においては、DNA 結合ドメインは 34 アミノ酸の 18 回繰返しで構成されている。
- 3 CRISPR/Cas9 においては、DNA 結合ドメインは二本鎖の DNA で構成されている。
- 4 CRISPR/Cas9 などのゲノム編集技術で DNA の配列に人為的な変更を加えた作物は、現行（2017 年 10 月現在）の日本の法律では、遺伝子組換え作物の表記をする義務が設けられていない。

問 14

脊椎動物の胚発生について、器官や組織と、その由来する胚葉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。

- 1 内胚葉：筋肉、神経
- 2 中胚葉：骨格、腎臓
- 3 外胚葉：脊髄、表皮
- 4 外胚葉：水晶体、爪

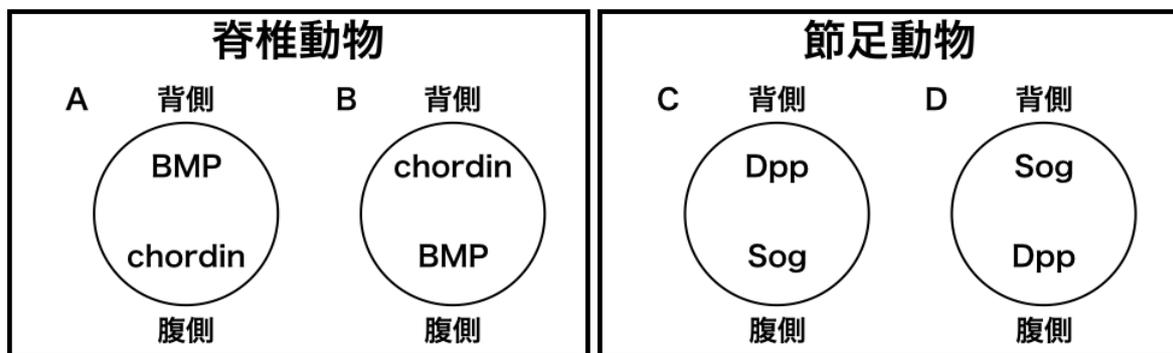
問 15

iPS 細胞(iPSC)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 iPS 細胞は「induced pluripotent stem cell」の略で、様々な組織や臓器の細胞に分化する能力とほぼ無限に増殖する能力をもつ。
- 2 iPS 細胞は山中伸弥教授らのグループによって世界で初めて作製され、山中教授はこの功績により 2012 年のノーベル生理学・医学賞を受賞した。
- 3 iPS 細胞は発生初期の胚盤胞から内部細胞塊を取り出し、それを培養することによって作製される。
- 4 *Oct3/4*, *Sox2*, *Klf4*, *c-Myc* など数種の転写因子遺伝子を体細胞に導入することでリプログラミングが起き、多能性幹細胞となる。

問 16

脊椎動物と節足動物では、前者は背中側に中枢神経系が、後者は腹側に神経系があることから、背腹軸が逆転していることが古くから知られていた。料理の下ごしらえでエビの背ワタを処理したことがあれば、容易に理解できるだろう。現在では、背腹軸の形成には脊椎動物と節足動物で同様の分子が関与しており、この軸が分子レベルでも逆転していることが知られている。節足動物であるハエの Short gastrulation(Sog)は脊椎動物の chordin のホモログであり、共に神経を誘導する。一方、ハエの Decapentaplegic(Dpp)は脊椎動物の BMP のホモログである。Sog/chordin は Dpp/BMP と拮抗し、その機能を阻害する。背腹軸の形成がおこる胞胚期において、これらのコードするタンパク質は脊椎動物と節足動物それぞれで、どのような局在となっていると考えられるか。以下の模式図の正しい組み合わせを選択肢の中から一つ選べ。



- 1 脊椎動物 : A    節足動物 : C
- 2 脊椎動物 : A    節足動物 : D
- 3 脊椎動物 : B    節足動物 : C
- 4 脊椎動物 : B    節足動物 : D

問 17

以下の細胞小器官のうち、動物細胞において核ゲノムの DNA とは別に独自のゲノム DNA を持つものを選択肢の中から一つ選べ。

- 1 ミトコンドリア
- 2 葉緑体
- 3 小胞体
- 4 ゴルジ体

問 18

動物界には有性生殖と無性生殖という二つの生殖様式が存在する。動物の生殖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 有性生殖では二倍体配偶子の接合によって二倍体の細胞である接合子が生じる。
- 2 無性生殖とは卵と精子の接合なしで起こる新個体の生産である。
- 3 単為生殖とは、無性生殖の一種で、卵が受精せずに発生することをいう。
- 4 動物の無性生殖の例として刺胞動物における出芽による増殖や、未受精卵が発生して雌親となる輪形動物の例が知られる。

問 19

クローン技術の人への応用に関する我が国（日本）における規制について、「ヒトに関するクローン技術等の規制に関する法律」があり、この規定に基づいて「特定胚の取扱いに関する指針（平成 21 年 5 月 20 日全部改正）」が文部科学省により作成されている。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 日本では、ヒトクローン胚を人以外の動物の体内に移植することは禁止されているが、ヒト胎内への移植は一定の条件の範囲で認められている。
- 2 日本では、ヒトと動物の間で卵子と精子をかけ合わせることにより作成された個体(ハイブリッド)を生む行為は禁じられている。
- 3 日本では、キメラ（ヒトと動物の細胞を融合させた個体）を作成する行為は禁じられている。
- 4 日本では、出産に結びつかないクローン研究は原則禁止とされているが、一定の条件の範囲内で認められている。

問 20

次の記述は日本の文部科学省/厚生労働省/経済産業省の三省によって共同で作成された、「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）の第一項の一部である。

「本指針は、遺伝情報が得られる等のヒトゲノム・遺伝子解析の特色を踏まえ、全てのヒトゲノム・遺伝子解析研究に適用され、研究現場で遵守されるべき倫理指針として策定されたものである。」

これに関連し、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 研究責任者は、ヒトゲノム・遺伝子解析研究の実施に当たって、あらかじめ研究計画書を作成し、研究を行う機関の長に許可を求めなければならない。
- 2 研究責任者は、個々の提供者の遺伝情報が明らかとなるヒトゲノム・遺伝子解析研究に関して、提供者が自らの遺伝情報の開示を希望していない場合には、原則として開示してはならない。
- 3 研究責任者は、提供者の同意がない場合には、提供者の遺伝情報を、提供者以外の人に対し、原則として開示してはならない。
- 4 研究責任者は、いかなる場合も、ヒトゲノム・遺伝子解析研究の業務を外部機関に委託してはならない。

問 21

以下の 2 つの式は、それぞれ、4 ビットの 2 の補数表現によるビット列で表現された整数間の演算を示している。これら 2 つの演算の結果の 2 の補数表現として正しいものを、選択肢の中から一つ選べ。なお、 $k$  ビットの 2 の補数表現  $C$  の値  $J$  は、 $C$  をそのまま 2 進整数だと考えた時の値を  $I$  とすると、 $C$  の先頭のビットが 0 であれば  $J = I$ 、 $C$  の先頭のビットが 1 であれば  $J = I - 2^k$  である。たとえば、 $k = 2$  の場合、 $C = "10"$  ならば  $J = -2$ 、 $C = "01"$  ならば  $J = 1$  である。

$$\begin{aligned} "0101" + "0001" &= \boxed{\quad ( a ) \quad} \\ "0011" - "1111" &= \boxed{\quad ( b ) \quad} \end{aligned}$$

- 1 (a) "0110" (b) "0100"
- 2 (a) "0110" (b) "1100"
- 3 (a) "0101" (b) "0100"
- 4 (a) "0101" (b) "1100"

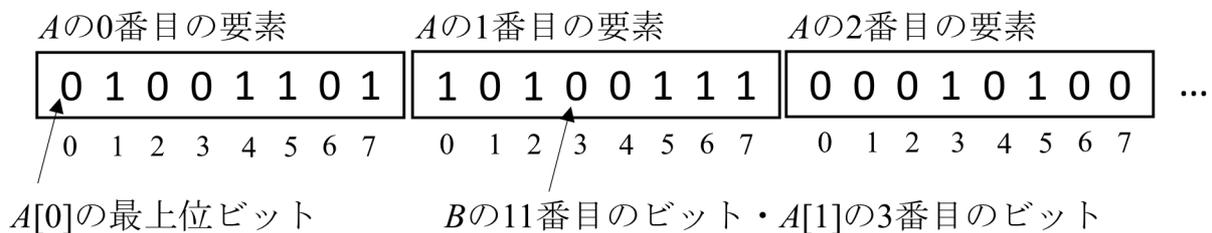
(問 22・問 23)

問 22

0 と 1 からなるビット列や DNA 配列など文字の種類数が少ない文字配列は、バイナリ配列として格納することでメモリやディスク領域を節約できることがある。

ここでは、長さ  $8n$  のビット列  $B$  を、8 ビット 2 進数を要素とする要素数  $n$  の配列  $A$  に保管することを考える。このとき、 $A$  の各要素内では  $B$  内の要素を上位ビットから順に並べるものとする。また、 $A$  内の要素、 $B$  内のビット、 $A$  の各要素内のビット、いずれもインデックスは 0 から数え始めるとする。すなわち、 $A$  の先頭の要素、 $B$  の先頭のビット、 $A$  の各要素それぞれの最上位ビットは、いずれもそれぞれの中での 0 番目と数える。

このとき、 $B$  の  $i$  番目のビットについて正しい記述を選択肢の中から一つ選べ。ただし、演算子  $/$  と  $\%$  は C 言語の仕様に則し、それぞれ整数除算における商と剰余を表すものとする。



- 1  $B$  の  $i$  番目のビットは  $A$  の  $i/8$  番目の要素の  $i\%8$  番目のビットに格納されている。
- 2  $B$  の  $i$  番目のビットは  $A$  の  $i/8$  番目の要素の  $(i\%8)+1$  番目のビットに格納されている。
- 3  $B$  の  $i$  番目のビットは  $A$  の  $(i/8)+1$  番目の要素の  $i\%8$  番目のビットに格納されている。
- 4  $B$  の  $i$  番目のビットは  $A$  の  $(i/8)+1$  番目の要素の  $(i\%8)+1$  番目のビットに格納されている。

問 23

問 22 の配列  $A$  について、 $j$  番目の要素の  $k$  番目のビットに格納されている値を計算する式として正しいものを選択肢の中から一つ選べ。ただしインデックスの数え方は前問と同様とする。なお、演算子  $\gg$  は C 言語の仕様に則し、右シフト演算を表すものとする。また、2 進数同士の論理和 (OR)、論理積 (AND) は、2 進での各桁ごとにそれぞれ論理和、論理積をとったものである。例えば、 $1110_{(2)} \text{ OR } 0101_{(2)} = 1111_{(2)}$ 、 $1110_{(2)} \text{ AND } 0101_{(2)} = 0100_{(2)}$  である。ここで数字列の末尾に付加した  $_{(2)}$  は、直前の数字列が 2 進数表現であることを示す。

- 1  $(A[j] \gg (8-k)) \text{ OR } 00000001_{(2)}$
- 2  $(A[j] \gg (7-k)) \text{ OR } 00000001_{(2)}$
- 3  $(A[j] \gg (8-k)) \text{ AND } 00000001_{(2)}$
- 4  $(A[j] \gg (7-k)) \text{ AND } 00000001_{(2)}$

問 24

プログラミング言語に関する以下の説明文中の( a )と( b )に入る用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。

プログラミング言語は大別するとコンパイラまたはインタプリタを利用するものがある。前者として、C,C++, ( a )などが挙げられる。後者としては、Perl,Python,( b )などが挙げられる。

- |   |             |            |
|---|-------------|------------|
| 1 | (a) Ruby    | (b) HTML   |
| 2 | (a) Ruby    | (b) Pascal |
| 3 | (a) FORTRAN | (b) Ruby   |
| 4 | (a) PHP     | (b) Ruby   |

問 25

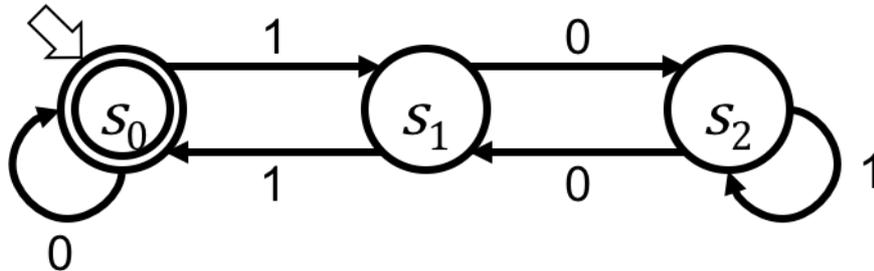
ゲノム情報など重要な個人情報や知財情報などを扱う際には、データや計算機の適切なセキュリティ保全・管理が必須であり、中でもデータ受け渡しや計算機の遠隔操作などの際のデータ通信の暗号化はきわめて重要である。そのようなデータ通信の際に用いられる暗号方式として、SSH などで用いられている公開鍵暗号方式が知られる。

この公開鍵暗号方式に関する下記の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 公開鍵暗号方式においては、受信者が秘密鍵と公開鍵の両方を作成し、送信者に公開鍵を配布する。
- 2 公開鍵暗号方式においては、公開鍵は平文の暗号化に用いられ、秘密鍵は暗号文の復号に用いられる。
- 3 公開鍵暗号方式においては、送信者以外の第三者に公開鍵が漏れないように注意する必要がある、「鍵配送問題」として知られている。
- 4 公開鍵暗号方式は共通鍵暗号方式と比べて復号に時間がかかるため、最初に共通鍵暗号の鍵の受け渡しをするためにだけ利用することが多い。

問 26

下の状態遷移図によって表現される有限状態オートマトンは、3つの状態  $s_0$ ,  $s_1$ ,  $s_2$  を持ち、 $s_0$  は初期状態であり、受理状態でもある。この有限状態オートマトンに関する記述で不適切なものを選択肢の中から一つ選べ。



- 1 このオートマトンは入力文字列 10111101 を受理するが 0110101 は受理しない。
- 2 このオートマトンが受理する文字列の集合は、正規表現  $(0 | 1(01^*0)^*1)^*$  と対応する。
- 3 状態と入力によっては次の遷移が一意に決定しないことがあるため、このオートマトンは非決定性有限状態オートマトンである。
- 4 このオートマトンが受理する文字列の集合と同一の文字列の集合を受理し、かつ、より状態数が少ない決定性有限状態オートマトンは存在しない。

(問 27・問 28)

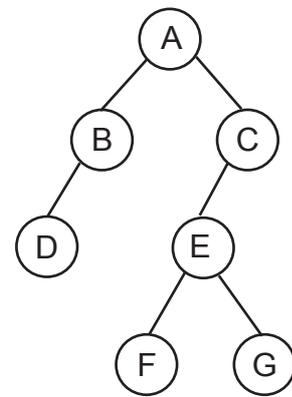
問 27

以下の擬似コードは木の巡回を行うプログラムである。このプログラムは、再帰呼び出しで巡回を実現している。

このプログラムでは、各ノードの行きがけで“(”を、帰りがけで)””を出力している。このプログラムで出力される括弧列は、**balanced parenthesis(BP)**表現とよばれ、木の構造を非常にコンパクトに表現することができる。なお、「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。

このとき、図の木に対して `TreeToParentheses (A)` を実行した際に得られる括弧列として正しいものを選択肢の中から一つ選べ。

```
TreeToParentheses (node v):  
begin  
  output("(");  
  foreach (c ∈ children(v)) do begin  
    TreeToParentheses (c);  
  end  
  output(")");  
end
```



- 1 ((()))(((())()))
- 2 (((())())
- 3 (((())())(())())
- 4 (((())())(())())

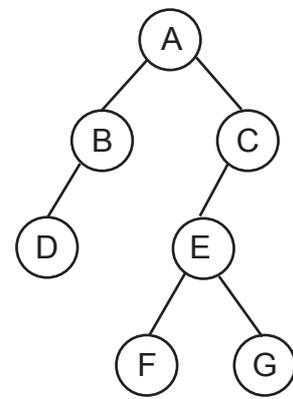
問 28

以下の擬似コードは、再帰呼び出しを用いず、スタック（擬似コード内の  $S$ ）を明示的に用いて木を巡回するプログラムである。

このプログラム  $\text{Traverse}(v)$  は、ノード  $v$  以下の子孫ノードを  $v$  も含めすべて巡回し、各ノードを  $\text{output}$  関数で出力する。なお、プログラム中の  $\text{push}$  と  $\text{pop}$  はスタック  $S$  に対する  $\text{push}$  操作と  $\text{pop}$  操作を表す。また、「 $\text{foreach}(w \in \text{children}(v)) \text{do} \dots$ 」は、前問同様、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。

このとき、図の木（前問と同じ木）に対して  $\text{Traverse}(A)$  を実行した際の出力として正しいものを選択肢の中から一つ選べ。なお、本問のプログラムの巡回順が問 27 のプログラムと同じとは必ずしも限らないことに注意せよ。

```
Traverse(node v):
begin
  push v to S ;
  while (S is not empty) do begin
    v ← pop a node from S ;
    output(v) ;
    foreach (w ∈ children(v)) do begin
      push w to S ;
    end
  end
end
end
```



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

### 問 29

ソートの計算量に関する **big-O** 記法を用い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。ただし、 $n$  は入力列の長さとする。

- 1 入力列中の要素同士の大小比較のみに基づいて要素の順位を決めるアルゴリズムで、任意の入力列を  $O(n)$  時間でソートできるものは存在しない。
- 2 基数ソート (ラディックスソート) は入力列中の要素同士の大小比較のみに基づくアルゴリズムではないため、適当な条件のもとで  $O(n)$  時間で入力列をソートすることができる。
- 3 バブルソートとマージソートは、どのような入力列に対しても  $O(n \log n)$  時間で動作する。
- 4 クイックソートは最悪の場合  $O(n^2)$  時間で動作し、入力データ以外に必要な作業領域は  $O(1)$  である。

### 問 30

ハッシュ法は、ハッシュ関数によってキーからハッシュ値を計算し、得られたハッシュ値を活用してハッシュ表中の格納場所を決定するアルゴリズムである。このとき、ハッシュ値が衝突を起こした時の代表的な対処法に連鎖ハッシュ法と開番地法がある。ハッシュ法についての以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 どのようなアルゴリズムを用いたハッシュ法でも、要素の探索時間は常に  $O(1)$  である。
- 2 ハッシュ法では、格納されるデータ数に対して十分大きなハッシュ表を用いれば、適切な仮定の下で、平均的に定数時間で要素を探索することができる。
- 3 開番地法では、ハッシュ値が衝突した場合、あいている別の番地を探す。
- 4 連鎖ハッシュ法では、ハッシュ値が衝突した場合、衝突を起こしたデータをポインタでつなぎ、リストで管理する。

### 問 31

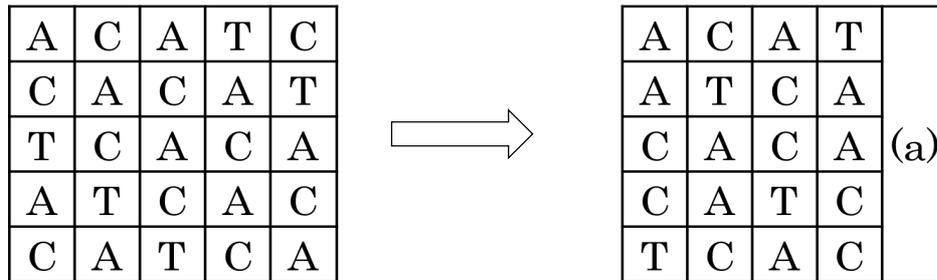
計算量の記述によく用いられる **big- $\theta$**  (ビッグ・シータ) 記法に関する以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。ただし記述中の関数はすべて、自然数を非負実数へ写す関数とする。

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

(問 32・問 33)

問 32

文字列  $s = ACATC$  の Burrows–Wheeler 変換を考える。まず、左下図のように、 $s$  に対して左から右に連続して 1 文字ずつ循環シフトして得られる文字列を行とする行列を作成する。ここで、 $s$  は行列の最初の行に対応する。この行列の行を右下図のように辞書式順序でソートして得られる行列の最後の列が、 $s$  に対する Burrows–Wheeler 変換を与える。



右上図中の ( a ) に当てはまる縦方向の文字列として正しいものを選択肢の中から一つ選べ。ここで、選択肢中の文字列の一番左の文字が ( a ) の一番上の文字に対応するものとする。

- 1 CTACA
- 2 CCTAA
- 3 CACAT
- 4 CATA C

問 33

以下の ( a )、( b )、( c ) に入れる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。

文字列  $s$  を Burrows–Wheeler 変換した文字列  $BW(s)$  の長さは  $s$  と ( a ) が、 $s$  が自然言語などのある構造をもったものであれば、 $BW(s)$  では ( b ) 文字が連続して出現する傾向にある。この性質を利用して  $s$  を ( c ) する技術が用いられている。

- 1 (a) 同じである      (b) 同じ      (c) 圧縮
- 2 (a) 異なる      (b) 異なる      (c) 伸長
- 3 (a) 同じである      (b) 異なる      (c) 圧縮
- 4 (a) 異なる      (b) 同じ      (c) 伸長

問 34

リレーショナルデータベースについての以下の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 リレーショナルデータベースは、データやデータ同士の関係をタプル（組）として表現し、タプルの集まりであるリレーションを保持する。
- 2 リレーショナルデータベースは、表として表現することが可能である。
- 3 リレーショナルデータベースを扱う標準的な言語として **SQL** が知られている。
- 4 データがドメイン制約（データを規定された範囲に収める制約）を満たしていれば、どのようなリレーショナルデータベースでも、その整合性が常に保たれる。

問 35

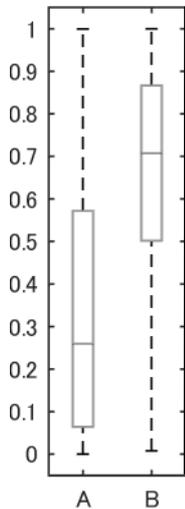
以下に示した 5 個の数値データが与えられたとする。このデータにおける 3 個の基本統計量（平均値、中央値、最頻値）の値として正しい組み合わせを選択肢の中から一つ選べ。ここで、中央値はメディアン、最頻値はモードとも呼ばれる。

データ : 4, 5, 4, 7, 6

選択肢	平均値	中央値	最頻値
1	5	5.5	5
2	5.2	5	4
3	5.2	5.5	4
4	5	5	5

問 36

以下の ( a ) ~ ( c ) に入れる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。



図は 2 標本 A,B の箱ひげ図を示している。標本 A および B の ( a ) は、それぞれ、0.064、0.502 であり、( b ) は、それぞれ、0.260、0.707 であり、( c ) は、それぞれ 0.571、0.867 であることをこの図は表している。

- |   |         |        |         |
|---|---------|--------|---------|
| 1 | (a) 5%点 | (b)中央値 | (c)95%点 |
| 2 | (a)25%点 | (b)中央値 | (c)75%点 |
| 3 | (a) 5%点 | (b)平均値 | (c)75%点 |
| 4 | (a)25%点 | (b)平均値 | (c)95%点 |

問 37

以下の二つの命題について、正しい説明を選択肢の中から一つ選べ。

ただし  $s$  と  $t$  は  $t < s$  を満たす正実数、 $m$  と  $n$  は  $n < m$  を満たす正整数とする。

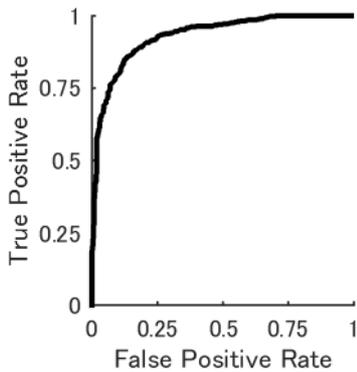
命題 a  $X$  が  $(0, s)$  上の一様分布に従うとき、 $(\frac{t}{s})X$  は  $(0, t)$  上の一様分布に従う。

命題 b  $X$  が  $\{0, 1, \dots, m - 1\}$  上の一様分布に従うとき、 $X \bmod n$  は  $\{0, 1, \dots, n - 1\}$  上の一様分布に従う。

- 1 命題 a も命題 b も正しい。
- 2 命題 a は正しいが命題 b は一般には正しくない。
- 3 命題 b は正しいが命題 a は一般には正しくない。
- 4 命題 a も命題 b も一般には正しくない。

問 38

2 値分類問題に対する予測手法の評価に ROC(Receiver Operating Characteristic)曲線がしばしば用いられる。ROC 曲線が下図のようであったとき、AUC(Area Under the Curve)としてもっとも適切な数値を選択肢の中から一つ選べ。



- 1 0.069
- 2 0.432
- 3 0.569
- 4 0.932

問 39

学習アルゴリズムなどにより構築した判別器の良さを確認するために交差検証法(cross-validation)が行われる。交差検証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。

- 1 交差検証法の目的は、未知の入力データに対する判別器の汎化誤差を推定することである。
- 2 単純なホールドアウト検証は訓練データとテストデータに分けた際の偶然による何らかの偏りによる影響を受けるが、交差検証法はそのような影響に対してより頑健である。
- 3 サンプルデータを  $n$  個に分け、そこから順に 1 つをテストのために確保し、残りを学習に用いる方法を  $n$ -fold 法とよぶ。
- 4 leave-one-out 法は、サンプルデータを分けることをせず、訓練データとテストデータに全く同一のデータ集合を用いることで、より厳密に判別器の良さを確認することができる検証法である。

#### 問 40

カーネル法は、特に SVM(Support Vector Machine)などと組み合わせてよく用いられるデータ解析手法であり、バイオインフォマティクスの諸問題でもよく活用されている。このカーネル法で用いられるカーネル関数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 カーネル関数とは、 $X$ を対象空間とすると、すべての  $x, y \in X$  について、特徴空間  $F$  への写像  $\varphi$  が存在し、 $K(x, y) = \varphi(x) \cdot \varphi(y)$  を満たすような関数  $K$  のことをいう。ここで、 $\varphi(x) \cdot \varphi(y)$  は  $\varphi(x)$  と  $\varphi(y)$  の内積を表す。
- 2  $x, y \in X = \{x_1, x_2, \dots, x_n\}$  に対して  $K(x, y)$  を  $X \times X$  から実数空間への対称関数とする。 $K(x, y)$  がカーネル関数となるための必要十分条件は、 $(i, j)$  成分が  $K_{ij} = K(x_i, x_j)$  で定義される  $n \times n$  行列  $K$  が半正定値となることである。
- 3  $K_1(x, y), K_2(x, y)$  をカーネル関数とすると、 $K_1(x, y) + K_2(x, y)$  もカーネル関数となる。
- 4 ガウスカーネルとは、連続する  $k$  文字の出現頻度を要素とする特徴ベクトルで定義されたカーネル関数のことである。

#### 問 41

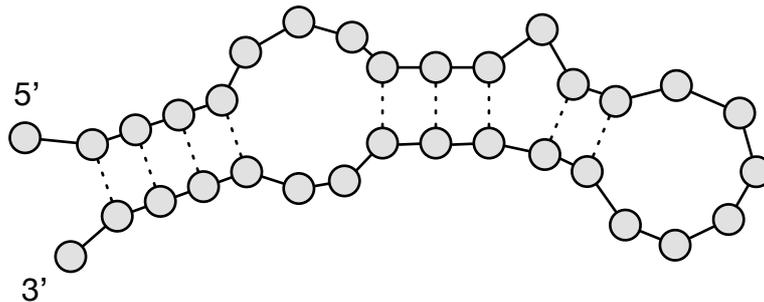
ショートリードのマッピング、すなわち、シーケンサにより得られた短いリードを参照ゲノム配列にマッピングする際には、リードとの類似度が同程度である箇所がゲノム上に複数見つかる場合がある。このような場合、多くのマッピングソフトウェアではリードを正確にゲノム上に位置づけることができないことを、マッピングクオリティ値などを用いて示すが、マッピング結果をどのように出力するかは用途によって切り替えられるようになっていることが多い。このような場合の出力方法として、通常考えられる用途の範囲においてもっとも不適切なものを選択肢の中から一つ選べ。

- 1 見つかった箇所からランダムに一箇所を選択し、リードのマッピングされた箇所として出力する。
- 2 見つかった箇所からゲノム座標が最も小さい箇所を選択し、リードのマッピングされた箇所として出力する。
- 3 見つかった箇所すべてを、リードのマッピングされた箇所として出力する。
- 4 見つかった箇所があまりに多い場合、リードのマッピングされた箇所として出力しない。

問 42

RNA の二次構造の表現方法のひとつに、括弧とピリオドを用いて表現する方法がある。この方法では、5' 末端側から塩基対を形成する場所に“(”をおき、対応する 3' 末端側に)””をおき、さらに、ループなどの塩基対を形成しない箇所は“.”とする。

下図は RNA 二次構造予測の結果であり、丸は塩基、実線はホスホジエステル結合、破線は塩基対を表す。上記の方法で下図の二次構造を表した表現として、もっとも適切なものを選択肢の中から一つ選べ。



- 1 .((((...(((.(.....))))).))))).
- 2 .((((...(((.(.....))))).))))).
- 3 .((((.(.(((.(.....))))).))))).
- 4 .((((...(((.(.....))))).))))).

問 43

次世代シーケンサ由来の配列データのアーカイブ化を行うにあたっては、これらの配列データが抱える様々な問題に対処することが必要となる。公的データベース NCBI の Sequence Read Archive(SRA) に関する以下の記述のうち、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 大容量のデータ転送に対応するために Aspera ソフトウェアが利用可能となっている。
- 2 すべての配列データは標準形式である FASTA フォーマットで公開され、FTP サイトからダウンロード可能である。
- 3 ヒトを対象としたデータなどは、公開に際してアクセス制限をかけることが可能である。
- 4 BioSample や BioProject というデータベースによって、配列の由来となったサンプルや研究プロジェクトの情報が管理され、アクセス可能となっている。

問 44

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 GEO : 遺伝子多型
- 2 JASPAR : 転写因子結合配列
- 3 Pfam : タンパク質ドメイン
- 4 UniProt : タンパク質配列

問 45

マルチプルアラインメント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 ガイドツリーに従って段階的にアラインメントを進める過程で、マルチプルアラインメント同士のアラインメントが行われる。
- 2 高速にマルチプルアラインメントを行うための手法として、高速フーリエ変換を応用したものがある。
- 3 マルチプルアラインメントの結果から、隠れマルコフモデルを用いた配列プロファイルを得ることができる。
- 4 マルチプルアラインメントの厳密な最適解は、動的計画法を用いることにより、配列の本数の二乗に比例した計算時間で求めることができる。

問 46

以下の ( a ) ~ ( d ) 内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。

真核生物のゲノム配列において、RNA ポリメラーゼ II による転写開始位置の ( a ) に見られる特徴として、( b ) 領域に含まれる ( c ) の存在がよく知られているが、遺伝子によってはこの特徴が見られないこともある。この他にも、転写制御に関わる様々な配列モチーフの存在が知られているものの、個々のモチーフ検索によって見つかる候補配列は、一般に大量の ( d ) を含んでおり、それらモチーフの組合せについてもそれほど明確かつ汎用的な規則が知られていないため、一般にゲノム配列のみから転写開始位置を正確に予測することは難しい。

- |   |       |           |              |        |
|---|-------|-----------|--------------|--------|
| 1 | (a)上流 | (b)プロモーター | (c)TATA ボックス | (d)偽陽性 |
| 2 | (a)上流 | (b)エンハンサー | (c)イニシエーター   | (d)偽陰性 |
| 3 | (a)下流 | (b)プロモーター | (c)TATA ボックス | (d)偽陰性 |
| 4 | (a)下流 | (b)エンハンサー | (c)イニシエーター   | (d)偽陽性 |

問 47

位置特異的スコア行列 PSSM(position-specific scoring matrix)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 PSSM におけるスコアとして、文字出現頻度をバックグラウンドと比べることで得られる対数尤度比がよく用いられる。
- 2 アミノ酸配列のパターンが可変長の挿入や欠失を含むとき、その表現方法として隠れマルコフモデルよりも PSSM の方が適している。
- 3 PSI-BLAST は、PSSM を用いることにより、BLAST に比べて、より類似度の低い相同配列まで検出することができる。
- 4 転写因子の結合部位を予測するために、PSSM を利用することができる。

問 48

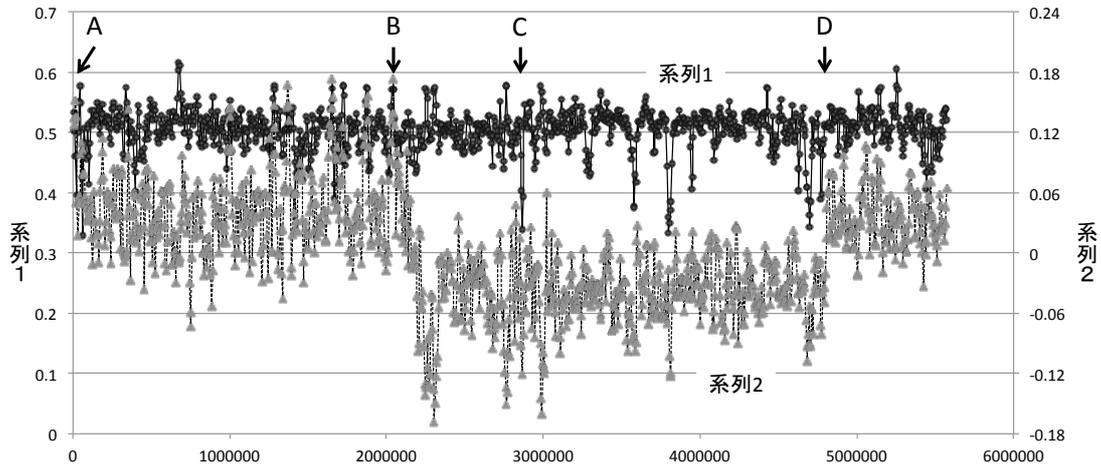
2つのゲノム間でオーソログな遺伝子対を見つける方法として、ゲノム中の遺伝子間で総当たりのホモロジー検索を行って、双方向ベストヒット基準に基づいて判定する方法がよく用いられる。この方法では、ゲノム A 中の遺伝子 a に対して、ゲノム B 中でもっとも高い類似性を示す遺伝子が b であり、かつ b から見てもゲノム A 中でもっとも高い類似性を示す遺伝子が a である場合に、a と b をオーソログと判定する。いま、問題を単純にするため、ゲノム A は遺伝子 a1 から a4 の 4 つ、ゲノム B は遺伝子 b1 から b5 の 5 つの遺伝子からなるとし、それらの間で総当たりのホモロジー検索を行った結果が以下の表のようであったとする（ただし、スコアは高いほど類似性が高いことを示す）。このとき、双方向ベストヒットによって抽出されるオーソログな遺伝子対はいくつあるか。選択肢の中から一つ選べ。

ゲノム A	ゲノム B	類似性スコア
a1	b1	280
a1	b2	160
a2	b3	70
a2	b4	80
a2	b5	110
a3	b1	110
a3	b2	120
a4	b1	160
a4	b2	220

- 1 1 対
- 2 2 対
- 3 3 対
- 4 4 対

問 49

以下の図は、ある細菌のゲノムについて、塩基組成に関する 2 つの量をゲノムに沿ってプロットしたものである。この図について説明した以下の文の空欄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。



図で系列 1 は ( ア ) を、系列 2 は ( イ ) を表しており、このゲノムの複製開始点は ( ウ ) の近傍にあると推定される。

選択肢	ア	イ	ウ
1	GC 含量	GC skew	A
2	GC skew	GC 含量	B
3	GC skew	GC 含量	C
4	GC 含量	GC skew	D

問 50

以下の配列アラインメントにおいて、これらの配列すべてにマッチするような（これら以外の配列にどれだけマッチするかは問わない）正規表現パターンを選択肢の中から一つ選べ。ただし、パターンはモチーフデータベース PROSITE の表記法に従っており、 $x(2)$  は任意のアミノ酸 2 個と、 $x(2,4)$  は任意のアミノ酸 2 個以上 4 個以下と、 $[FY]$  はアミノ酸 F または Y と、それぞれマッチすることを表す。また、アラインメント上部の \* はすべての配列において同一のアミノ酸をとる部位を、+ はそれ以外でアミノ酸の保存性が高い部位を示している。

```

*      *      +      +      *      *
CNI--CDRGFTTKGNLQHML--TH
CSE--CSRSFTHNSDLTAHMR--KH
CSFDGCEKVYNRPSLLQQHQ--SH
CEY--CGKRFHQKSDMKKHTY--IH
CPF--CAANYPTHFDLQKHLI-QEH
CEI--CDKTFATRLSLQDHLFAEQH
CFE--CGIGFDTAHRLEVHHD-ERH

```

- 1 C-x(4)-C-x(12)-H-x(5)-H
- 2 C-x(2,4)-C-x(3)-[FYW]-x(8)-H-x(2,5)-H
- 3 C-x(3,5)-C-x(3)-[FW]-x(5)-[LM]-x(2)-H-x(2,4)-H
- 4 C-x(2,4)-C-x(5)-[FY]-x(3)-[LVI]-x(2)-H-x(3,5)-H

(問 51・問 52)

以下の図は、動的計画法によって 2 本の配列の大域的アラインメントスコアを計算する途中過程を表しており、各セルの値は左上隅のセルからそのセルへ至る最適経路（スコアが最大となる経路）のスコアを表している。ただし、スコア計算では塩基の一致、不一致、ギャップの 3 状態のみを考慮する（開始、延長、末端の区別はしない）ものとする。これについて以下の問 51 および問 52 に答えよ。

		C	A	G	T
A	0	-2	-4	-6	-8
G	-2	-1	-1	-3	-5
C	-4	-3	-2		
			(※)		

問 51

このアラインメント計算における一致、不一致、ギャップに対するスコア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。

選択肢	一致	不一致	ギャップ
1	2	0	-1
2	1	-1	-2
3	1	0	-1
4	2	-1	-2

問 52

このスコア体系に従って計算を続けたとき、(※)のセルに入るスコアの値はいくらになるか。選択肢の中から一つ選べ。

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

問 53

PDBにおける主要なデータフォーマットである mmCIF では、データをキーとバリューの組み合わせとして記述している。キーはカテゴリーグループ、カテゴリ、アイテムという階層構造をとっている。例えば、

<code>_em_assembly.id</code>	1
<code>_em_assembly.entry_id</code>	5FLM

という内容であれば、カテゴリーグループ `em` に属するカテゴリ `em_assembly` における、アイテム `id` には“1”というバリューが設定されている。また同じカテゴリのアイテム `entry_id` には“5FLM”というバリューが設定されている。

以下はタンパク質立体構造データ(PDB:5FLM)の一部である。このデータ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。

```

_pdbx_database_related.db_name      EMDB
_pdbx_database_related.db_id        EMD-3218
_pdbx_database_related.content_type  'associated EM volume'
#
(中略)
_em_2d_projection_selection.entry_id  5FLM
_em_2d_projection_selection.software_name  'RELION 1.3'
_em_2d_projection_selection.num_particles  264134
#
(中略)
#
_em_assembly.id                      1
_em_assembly.entry_id                5FLM
_em_assembly.aggregation_state       PARTICLE
_em_assembly.name                    'BOVINE POL II ELONGATION COMPLEX'
_em_assembly.details                 ?
#
_em_buffer.id                        1
_em_buffer.name                      '150 MM NAACL, 5 MM HEPES, 0.01 MM ZNCL2, 10 MM DTT'
#
_em_detector.entry_id                5FLM
_em_detector.id                      1
_em_detector.type                    'GATAN K2 SUMMIT'
#
_em_experiment.entry_id              5FLM
_em_experiment.reconstruction_method 'SINGLE PARTICLE'
_em_experiment.specimen_type         'VITREOUS ICE'
#
_em_image_scans.entry_id             5FLM
_em_image_scans.id                   1
_em_image_scans.number_digital_images 1172

```

- 1 このデータは電子顕微鏡で解析されたウシ RNA ポリメラーゼ II 伸張複合体の立体構造である。
- 2 1,172 個の粒子についての測定データから立体構造を再構成した。
- 3 二次元画像からの立体構造の再構成には RELION というソフトウェアを用いた。
- 4 原子座標の元となった電子密度マップは EMDDataBank のエントリ EMD-3218 として登録されている。

問 54

以下の文中の ( a ) ~ ( c ) 内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。

分子シミュレーションにおける 2 原子間のポテンシャルエネルギーの計算には下記の 2 つの式がよく用いられる。

$$V_{vdW}(r) = A \left\{ \left( \frac{r_0}{r} \right)^{12} - 2 \left( \frac{r_0}{r} \right)^6 \right\}$$

$$V_{ele}(r) = B \frac{q_1 q_2}{r}$$

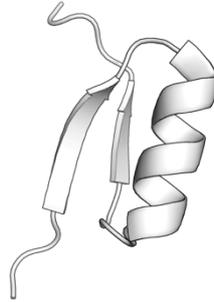
$V_{vdW}$ ,  $V_{ele}$  はそれぞれファンデルワールス相互作用、静電相互作用のポテンシャルエネルギーである。 $A$ ,  $B$ ,  $r_0$ ,  $q_1$ ,  $q_2$  はパラメータであり、 $r$  は 2 原子間の距離を示している。系全体のポテンシャルエネルギーを求めるには、単純に考えるとすべての原子ペアについて  $V_{vdW}$ ,  $V_{ele}$  を求めて総和を取る必要があるため、計算量は系を構成する原子数  $N$  に依存し、( a ) となる。しかし ( b ) は距離の増加に伴い速やかに減衰して通常では  $10 \sim 12 \text{ \AA}$  程度で十分にゼロに近づくため、遠い距離にある原子ペアを無視することで計算量を削減できる。もう一方の式は距離による減衰が遅いため、原子の分布が ( c ) であると仮定して無限遠の相互作用まで考慮する計算方法がよく用いられる。

- |   |              |               |         |
|---|--------------|---------------|---------|
| 1 | (a) $O(N)$   | (b) $V_{ele}$ | (c) 周期的 |
| 2 | (a) $O(N^2)$ | (b) $V_{vdW}$ | (c) 周期的 |
| 3 | (a) $O(N)$   | (b) $V_{vdW}$ | (c) 再帰的 |
| 4 | (a) $O(N^2)$ | (b) $V_{ele}$ | (c) 再帰的 |

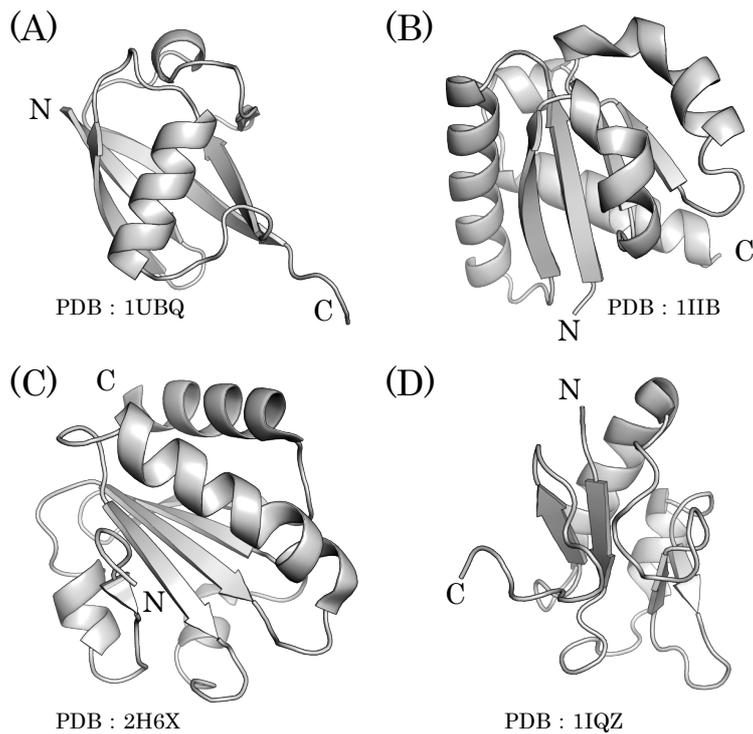


問 56

以下の図は、 $\beta-\alpha-\beta$  と呼ばれる超二次構造であり、 $\alpha/\beta$  クラスのタンパク質に含まれる特徴的な部分構造である。



以下の図(A)~(D)のタンパク質のうち、 $\beta-\alpha-\beta$  超二次構造を含む構造として、もっとも適切なものを選択肢の中から一つ選べ。



- 1 (B)
- 2 (A)と(B)
- 3 (B)と(C)
- 4 (C)と(D)

### 問 57

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、ベクトル表現を用いて  $\mathbf{x} = (x_1, x_2, \dots, x_i, \dots, x_n)$  と表現される。ここで  $\mathbf{x}$  の各要素  $x_i \in \{0,1\}$  は、タンパク質の  $i$  番目の残基と低分子  $X$  との相互作用の有無を意味する。

いま、興味のあるタンパク質と低分子の複合体について、相互作用プロファイルを用いてデータベースに存在する複合体情報の中から類似した相互作用を有する複合体を検索したい。相互作用の類似の度合いは相互作用プロファイル同士の Tanimoto 係数（Jaccard 係数）によって計算できる。

あるタンパク質と低分子  $A, B, C, D, Q$  の相互作用プロファイルが以下に示すベクトル  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{q}$  で表される時、Tanimoto 係数に基づいて低分子  $Q$  と最も類似した相互作用を有する低分子を選択肢の中から一つ選べ。ただし、Tanimoto 係数は  $\mathbf{x} \cdot \mathbf{y} / (\mathbf{x} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} - \mathbf{x} \cdot \mathbf{y})$  で計算される値であり、大きい値であるほど 2 つのベクトルが類似していることを表す。 $\cdot$  はベクトルの内積を表す。

低分子 $A$ :	$\mathbf{a} = (1, 1, 0, 0, 1, 1, 1)$
低分子 $B$ :	$\mathbf{b} = (1, 0, 1, 1, 0, 0, 1)$
低分子 $C$ :	$\mathbf{c} = (0, 1, 0, 0, 1, 0, 1)$
低分子 $D$ :	$\mathbf{d} = (1, 1, 1, 1, 1, 0, 1)$
低分子 $Q$ :	$\mathbf{q} = (1, 0, 0, 0, 1, 0, 1)$

- 1 低分子  $A$
- 2 低分子  $B$
- 3 低分子  $C$
- 4 低分子  $D$

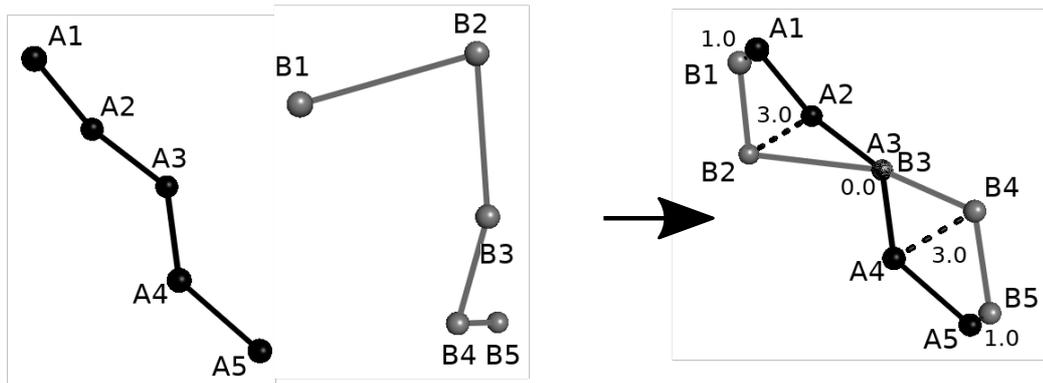
### 問 58

タンパク質をその存在する環境の違いにより水溶性タンパク質と膜タンパク質に大別して考える。両者における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 水溶性タンパク質も膜タンパク質も、内部に埋もれたアミノ酸と外部に露出したアミノ酸の性質は大きく異なる。
- 2 水溶性タンパク質も膜タンパク質も、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進化的保存度が高い傾向がある。
- 3 水溶性タンパク質の分子表面において、機能的制約がかかっているアミノ酸は進化的保存度が高い傾向にある。
- 4 膜タンパク質の分子内に埋もれたアミノ酸は疎水的であり、脂質二重層と接するアミノ酸は親水的である。

問 59

それぞれ 5 アミノ酸（アミノ酸 A1～A5 とアミノ酸 B1～B5）からなる 2 つのペプチドの立体構造を、下図に示すように重ね合わせる。対応するアミノ酸の C $\alpha$  原子間距離は以下の通りであった。この構造の重ね合わせにおける C $\alpha$  原子間 RMSD (Root Mean Square Deviation) 値として最も近い値の番号を選択肢の中から一つ選べ。

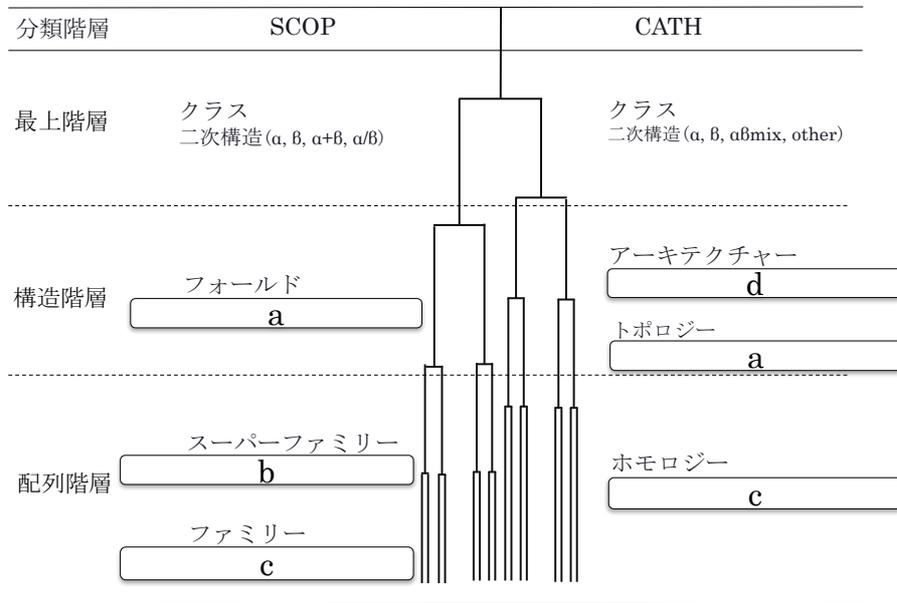


A1 - B1 間	1.0 Å
A2 - B2 間	3.0 Å
A3 - B3 間	0.0 Å
A4 - B4 間	3.0 Å
A5 - B5 間	1.0 Å

- 1 0.9 Å
- 2 2.0 Å
- 3 4.5 Å
- 4 8.0 Å

問 60

下図は SCOP と CATH のフォールド分類階層を図示したものである。各分類階層についての説明 a~d に入る説明とし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。



選択肢	a	b	c	d
1	配列順を考慮した二次構造配置	フォールドや機能類似性をもとに相同と推定される	高い配列類似性をもとに相同と判定される	配列順を考慮しない二次構造配置
2	配列順を考慮しない二次構造配置	フォールドや機能類似性をもとに相同と推定される	高い配列類似性をもとに相同と判定される	配列順を考慮した二次構造配置
3	配列順を考慮した二次構造配置	機能類似タンパク質	フォールドや機能類似性をもとに相同と推定される	配列順を考慮しない二次構造配置
4	配列順を考慮しない二次構造配置	機能類似タンパク質	フォールドや機能類似性をもとに相同と推定される	配列順を考慮した二次構造配置

問 61

量的形質(Quantitative Trait)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 穀物の品質、動物の肉質、身長、体重などのように、数値で連続的に表される形質のことである。
- 2 遺伝子だけでなく環境要因の影響に支配されることが多い形質で、メンデルの法則にしたがわずに遺伝する。
- 3 QTL (Quantitative Trait Locus : 量的形質座位) 解析とは、注目する形質と遺伝子マーカーとの連鎖をみることで QTL の場所、数、効果の大きさを検出する統計学的手法である。
- 4 ヒトにおいては家系解析を行うことで、交配実験が可能な動植物においては交配実験を重ねることで QTL を検出する。

問 62

ゲノムワイド関連解析(Genome Wide Association Study:GWAS)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 ある疾患や形質に関連する遺伝子マーカー（主として SNP）を、全ゲノムを対象に網羅的に検索する方法である。
- 2 疾患群と対照群から SNP を検出して、ある座位で特定の SNP が疾患群で有意に多くみられる時、その SNP 近傍のゲノム領域が疾患に関連すると考える。
- 3 結果の解析によく利用される指標には、 $-\log_{10}p$  値やオッズ比がある。
- 4 GWAS においてメタアナリシスを行うと、これまで報告された研究すべてを使って大規模な解析が行えるため、データソースに関するバイアスが排除できる。

問 63

タンパク質をコードしている遺伝子の進化の過程における塩基置換は、同義置換と非同義置換に分類される。同義置換と非同義置換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。同義置換率は同義サイトあたりの置換数、非同義置換率は非同義サイトあたりの置換数である。

- 1 同義置換はアミノ酸を変化させない置換で、非同義置換はアミノ酸を変化させる置換である。
- 2 大部分の遺伝子では、同義置換率は非同義置換率よりも小さい。
- 3 ウイルスの表面抗原の抗体によって認識される部位では、抗体による認識を避けるためアミノ酸置換が加速されることがあり、非同義置換率が大きくなる。
- 4 マウスとヒトのオーソログ遺伝子を比較する場合、同義置換率は遺伝子によらずほぼ一定の値になるが、非同義置換率はコードされる遺伝子によって大きく異なる。

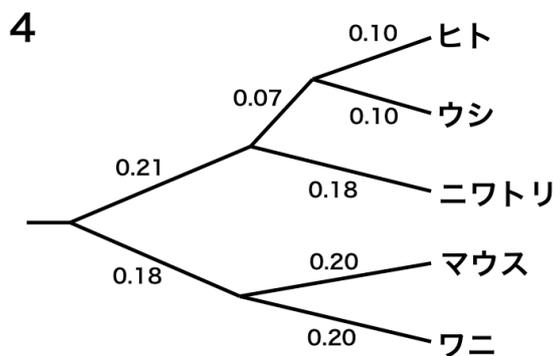
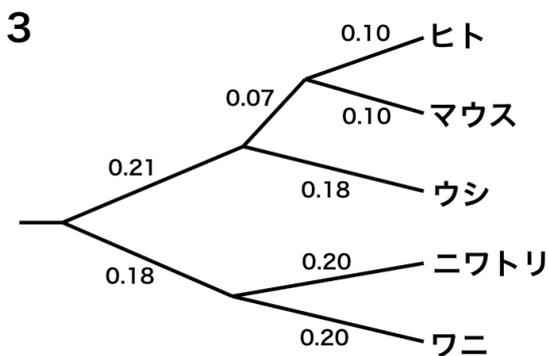
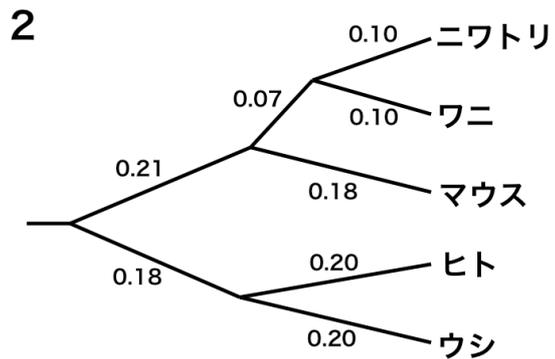
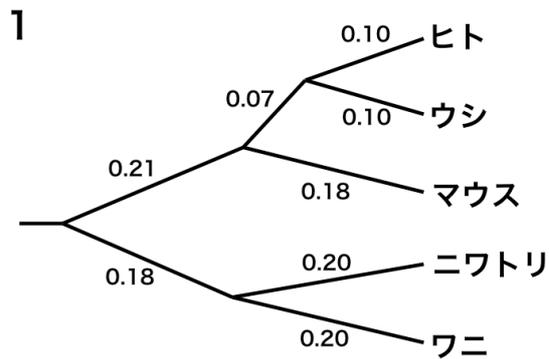
(問 64・問 65・問 66)

以下の表は 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) を求めたものである (自明である対角および右上成分は表示していない)。この表に基づいて以下の問 64、問 65、および問 66 に答えよ。

	ヒト	ウシ	マウス	ニワトリ	ワニ
ヒト					
ウシ	0.2				
マウス	0.3	0.4			
ニワトリ	0.7	0.7	0.9		
ワニ	0.7	0.7	0.9	0.4	

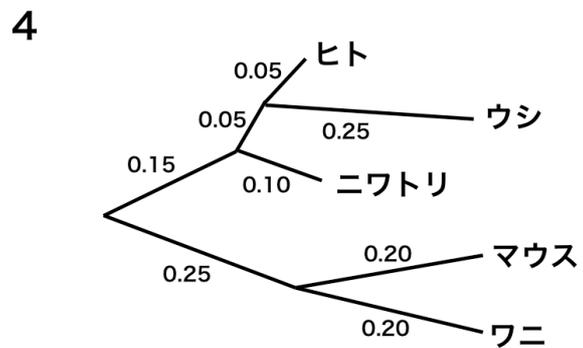
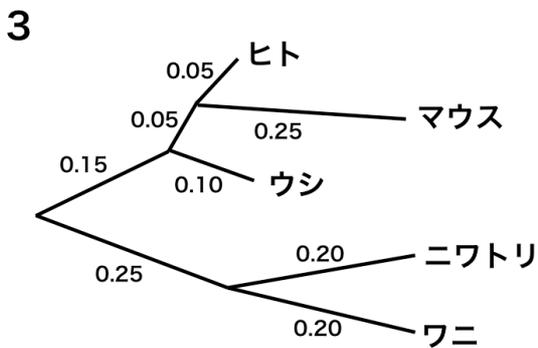
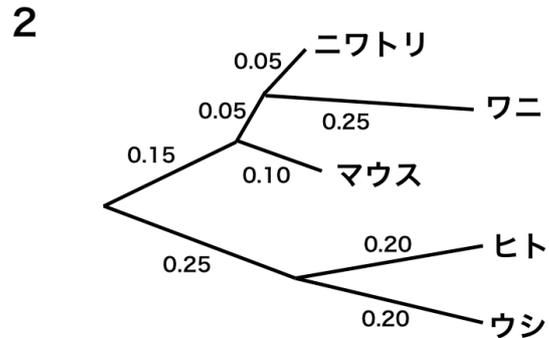
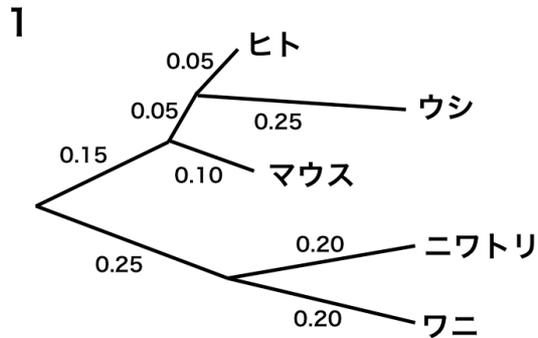
問 64

上記の進化距離から非加重平均距離法 (UPGMA 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。系統樹の各枝に示された数字は、その枝の長さ (進化距離) である。



問 65

問 64 の表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。系統樹の各枝に示された数字は、その枝の長さ (進化距離) である。



問 66

問 64 と問 65 で求めた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 系統樹作成法には大きく分けて、非加重平均距離法 (UPGMA 法) や近隣結合法 (NJ 法) のように距離行列を用いる方法と、最尤法 (ML 法) や最大節約法 (MP 法) のように配列データを直接用いる方法がある。
- 2 UPGMA 法では有根系統樹が得られるが、NJ 法で得られるのは無根系統樹である。
- 3 2つの方法で求めた系統樹のトポロジーおよび枝長に違いがある主な要因は、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に違いがあるのは、マウス由来の遺伝子が水平伝播によってもたらされたためである。

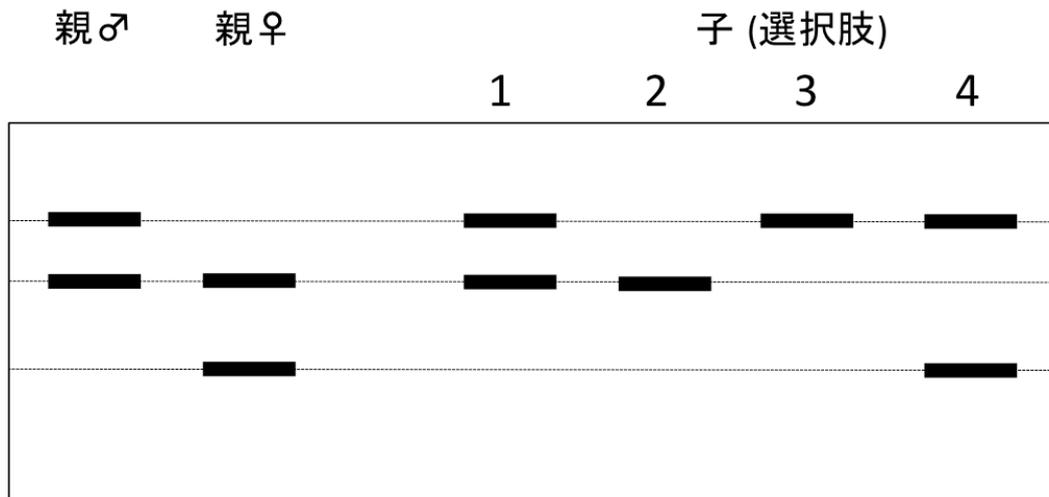
問 67

遺伝子連鎖解析の結果、遺伝子座 A と B の間で 100 回に 1 回の割合で組換えが観察された。この遺伝子座の間の遺伝学的距離として、もっとも適切なものを選択肢の中から一つ選べ。

- 1 0.01 cM
- 2 0.10 cM
- 3 1.00 cM
- 4 10.0 cM

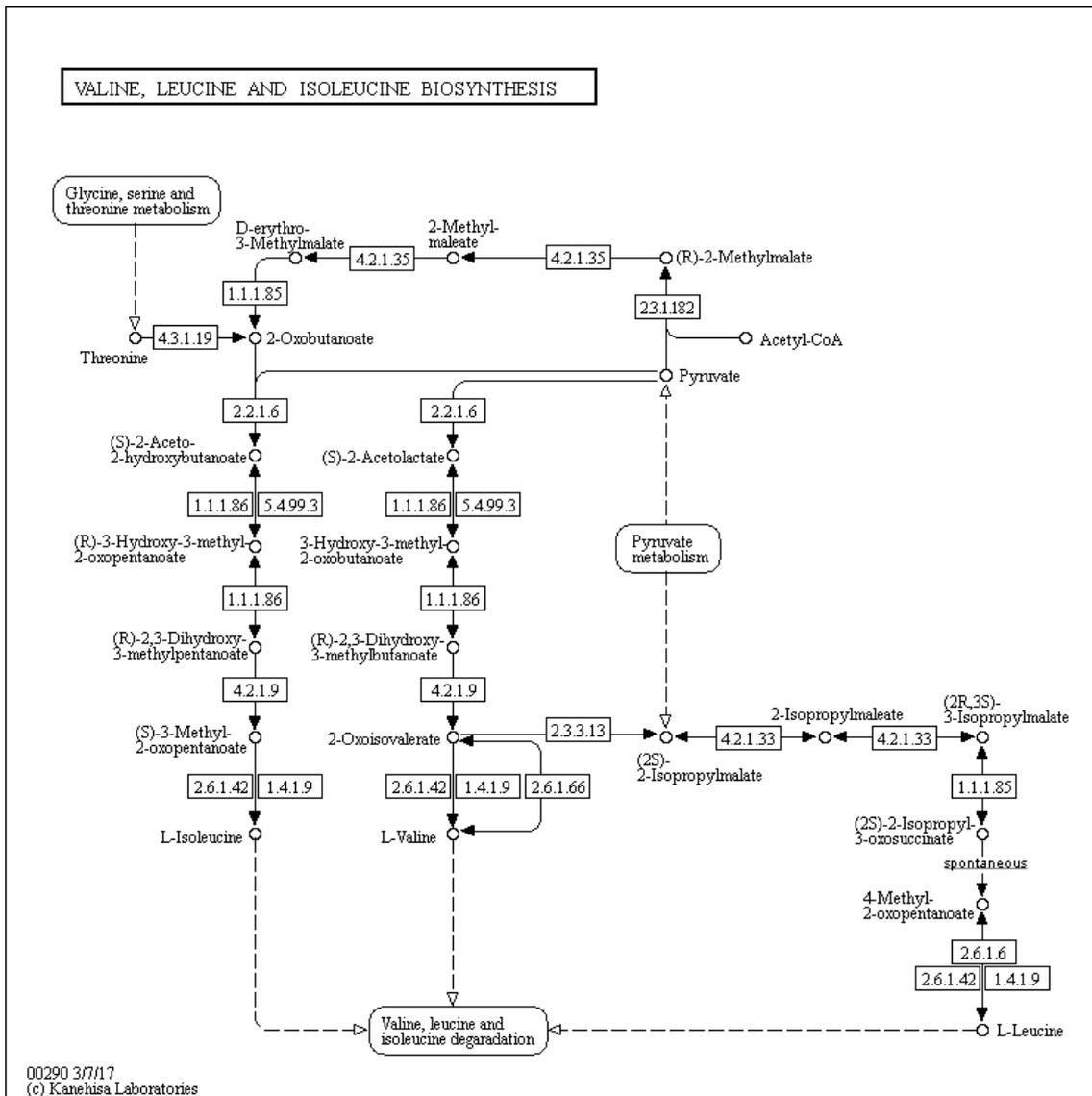
問 68

飼育しているマウスの親子関係が不明になったため、マイクロサテライト領域の多型を利用して DNA 鑑定を行った。下の図は、常染色体上の特定座位のマイクロサテライトを増幅するプライマーを用いて、両親（親♂および親♀）およびその両親の子供であることを検査するマウス（子 1~4）のゲノムに対して PCR 増幅を行い、増幅産物を電気泳動で分離したゲルの模式図である。この両親の子供でない可能性がもっとも高い個体を選択肢の中から一つ選べ。ただし、生殖過程における染色体の転座・転移・重複、およびマイクロサテライト反復数の変化は起こっていないものと仮定する。



(問 69・問 70)

下記の図は KEGG のパスウェイマップである。これを見て以下の問 69 および問 70 に答えよ。



問 69

上のパスウェイマップに関して述べた次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 図中の四角は酵素タンパク質を意味し、中の数字は対応する遺伝子番号である。
- 2 図中の○は代謝化合物を意味している。
- 3 化合物によっては、酵素タンパク質の反応を経ずに別の化合物になることもある。
- 4 化合物によっては、複数の酵素タンパク質により代謝されることもある。

問 70

問 69 のパスウェイマップの理解に有用なオーミクス解析に使われる技術について述べた次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 酵素の機能に影響を与えることがある遺伝子の変異の影響を見るためには、ゲノム変異解析が有用である。
- 2 代謝化合物の定量は、次世代シーケンサによって大幅に加速された。
- 3 代謝に係わる酵素タンパク質の量は、質量分析装置により推定することができる。
- 4 酵素タンパク質の機能解析には、タンパク質の立体構造解析が有用である。

問 71

RNA-seq やマイクロアレイ等の遺伝子発現量データの解析について述べた次の記述のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 RNA-seq ではマイクロアレイと比べて、未知の転写産物や、既知の遺伝子における変異を検出できるという利点がある。
- 2 多サンプルの発現量データに基づいて、サンプルだけでなく遺伝子も分類することができる。
- 3 RNA-seq ではマイクロアレイと異なり、発現の絶対量が測定できるため、階層型クラスタリングにはユークリッド距離を利用することができる。
- 4 2 群のサンプル間で発現量に差がある遺伝子が一定数見つかった場合、遺伝子に付与された遺伝子オントロジー語句(GO term)の統計的な偏りを解析し、遺伝子群の特徴を推定することが有効である。

問 72

次世代シーケンサとその使い分け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 Illumina 社の HiSeq シリーズは、1 ランあたりのリード数が多く、個人ゲノムの決定に使われている。
- 2 Illumina 社の MiSeq シリーズは、1 ランあたりのリード数が HiSeq より多く、リード長も長いので、メタゲノム解析によく使われる。
- 3 PacBio 社の RS シリーズは、リード長が長いので、ゲノム上のリピート配列をまたいで読むことができる可能性が高くなる。
- 4 Oxford Nanopore 社の MinION においては、ナノスケールの穴に DNA を通し、その際のイオン電流の変化で塩基を読み出す。ナノスケールの穴を実現するのに  $\alpha$ -hemolysin という 7 量体のタンパク質が利用されている。

問 73

次世代シーケンサのデータ解析では、解読された多数のリードと、ヒトゲノムや遺伝子配列データベースなどの参照配列を比較・照合する解析が行われることがある。以下のうち、参照配列やデータベースに基づく解析ではないものを選択肢の中から一つ選べ。

- 1 全ゲノム解析による、個人に特異的な一塩基置換 (SNP) の検出
- 2 糞便の 16S rRNA 配列決定とデータベース照合による腸内細菌叢の推定
- 3 長鎖シーケンサデータの *de novo* アセンブリによる全ゲノム推定
- 4 転写因子結合部位やヒストン修飾部位を決めるための ChIP-seq 解析

問 74

4 つの遺伝子(A,B,C,D)から構成される遺伝子ネットワークの制御構造を推定するために、それぞれの遺伝子を破壊する実験を行った。実験と結果は以下の通りである。この実験の評価として、もっとも不適切なものを選択肢の中から一つ選べ。

<実験と結果>

- 遺伝子 A を破壊すると、遺伝子 B、C、D の発現量が減少した。
- 遺伝子 B を破壊すると、遺伝子 A の発現量が増加し、遺伝子 C、D の発現量が減少した。
- 遺伝子 C を破壊すると、遺伝子 A、B、D の発現量が増加した。
- 遺伝子 D を破壊しても、他の遺伝子の発現量は変化しなかった。

- 1 遺伝子 A は、遺伝子 B の発現量を増加させる。
- 2 遺伝子 B は、遺伝子 C の発現量を増加させる。
- 3 遺伝子 C は、遺伝子 A の発現量を減少させる。
- 4 遺伝子 A、B、C は、正のフィードバックループを形成している。

問 75

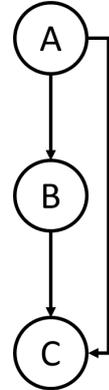
次世代シーケンサを用いた解析手法の説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 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。
- 2 DNase-Seq 解析は、オープンクロマチン領域をゲノムワイドに決定することができる。
- 3 RNA-Seq 解析は、ゲノム DNA のメチル化状態を明らかにすることができる。
- 4 Hi-C 解析は、核内におけるゲノム DNA の三次元構造を明らかにすることができる。

問 76

A,B,Cの遺伝子からなる遺伝子制御ネットワークには、理論的には図で示したような構造が存在しうる。A→BはAがBを制御することを意味する。このネットワークの説明として、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 Aの発現は最終的に自分自身を抑制する方向に働く。
- 2 この制御はフィードフォワードループと呼ばれる。
- 3 CはAとBの両方から制御を受けている。
- 4 この制御構造は遺伝子制御ネットワークに広く存在することが知られている。



問 77

細胞内ネットワークの推定方法について、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 個別の要素を積み上げて作るボトムアップ型は網羅性を上げるのが困難であるが個別に相互作用を確認するため信頼性が高く、大量データから推定を行うトップダウン型では網羅性は高いが偽陽性も多い。
- 2 複数の遺伝子について、異なる実験条件下で発現量を測定しその相関係数を計算することで、遺伝子の制御関係を推定することができる。
- 3 微分方程式を用いてネットワークの動的挙動とネットワーク構造をモデル化することができる。
- 4 ベイジアンネットワークによる推定においては、各々の遺伝子の発現量を確率変数とみなし、細胞の状態をそれら確率変数の事前確率分布として表す。

問 78

時間  $x$  の関数である変数  $y_1, y_2$  からなるシステムの連立微分方程式が

$$\begin{aligned}\frac{dy_1}{dx} &= -y_1 + 2y_2 \\ \frac{dy_2}{dx} &= 6y_1 + 3y_2\end{aligned}$$

のように表される。行列

$$\begin{pmatrix} -1 & 2 \\ 6 & 3 \end{pmatrix}$$

の固有値を用いて、システムの安定性を論じることができるが、平衡状態（原点）におけるシステムの安定性に関する記述でもっとも適切なものを選択肢の中から一つ選べ。安定とは、システムに摂動が与えられた場合、すみやかに元の平衡状態にもどれることをいう。

- 1 固有値は  $5, -3$  であるので、システムは安定である。
- 2 固有値は  $5, -3$  であるので、システムは不安定である。
- 3 固有値は  $-5, 3$  であるので、システムは安定である。
- 4 固有値は  $-5, 3$  であるので、システムは不安定である。

問 79

ゲノム・トランスクリプトーム・プロテオームなどの生体内高分子の解析に加えて、近年、糖・アミノ酸・有機酸・脂質などの生体内低分子化合物を網羅的に解析するメタボローム解析が、より表現型に直結する解析手法として注目されている。質量分析器によるメタボローム解析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 化合物の同定は分離の特性と推定質量をもとに、標準化合物ライブラリーを検索することで行う。
- 2 単一の分離技術で全代謝物質を精度良く分離するのは、代謝物の濃度範囲が広く種類も多いため困難である。
- 3 網羅的解析による物質同定は、液体クロマトグラフィーと質量分析器による分離の後にX線による構造解析を組み合わせる。
- 4 同位体標識された化合物の変化を時系列測定することにより、代謝流束（フラックス）解析が可能である。

問 80

トランスクリプトーム解析やメタボローム解析等の定量的な実験では、データに内在するばらつきの影響を評価するために実験を反復することが多い。実験の反復には大きく分けて生物学的反復(**biological replicates**)と技術的反復(**technical replicates**)がある。これらについて述べた以下の文章のうち、もっとも不適切なものを選択肢の中から一つ選べ。

- 1 生物学的反復とは、生物サンプルの調整から反復することである。
- 2 技術的反復とは、同じ生物サンプルから得られたデータを、異なる手法で反復解析することである。
- 3 測定に高いコストがかかる場合には、生物サンプルの調整は複数回行い測定試料をプールするが、測定は1回ですませ、生物学的反復とすることもある。
- 4 高精度な測定のためには、技術的反復の回数だけでなく生物学的反復の回数を増やすことが重要である。

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