

--	--	--	--	--	--

東京大学 大学院新領域創成科学研究科 情報生命科学専攻

Department of Computational Biology, Graduate School of Frontier Sciences

The University of Tokyo

平成 23(2011)年度

2011 School Year

大学院入学試験問題 修士・博士後期課程

Graduate School Entrance Examination Problem Booklet, Master's and Doctoral Course

専 門 科 目

Specialties

平成 22 年 8 月 24 日(火)

Tuesday, August 24, 2010

13:30~15:30

注意事項 Instructions

1. 試験開始の合図があるまで、この冊子を開いてはいけません。
Do not open this problem booklet until the start of examination is announced.
2. 本冊子の総ページ数は 14 ページです。落丁、乱丁、印刷不鮮明な箇所などがあつた場合には申し出ること。
This problem booklet consists of 14 pages. If you find missing, misplaced, and/or unclearly printed pages, ask the examiner.
3. 解答には必ず黒色鉛筆(または黒色シャープペンシル)を使用しなさい。
Use black pencils (or mechanical pencils) to answer the problems.
4. 問題は 8 題出題されます。問題 1~8 から選択した合計4問に解答しなさい。ただし、問題 1~8 は同配点です。
There are eight problems (Problem 1 to 8). Answer four problems out of the eight problems. Note that Problem 1 to 8 are equally weighted.
5. 解答用紙は計4枚配られます。各問題に必ず1枚の解答用紙を使用しなさい。解答用紙に書ききれない場合は、裏面にわたつてもよい。
You are given four answer sheets. You must use a separate answer sheet for each problem. You may continue to write your answer on the back of the answer sheet if you cannot conclude it on the front.
6. 解答は日本語または英語で記入しなさい。
Answers should be written in Japanese or English.
7. 解答用紙の指定された箇所に、受験番号と選択した問題番号を記入しなさい。問題冊子にも受験番号を記入しなさい。
Fill the designated blanks at the top of each answer sheet with your examinee number and the problem number you are to answer. Fill the designated blanks at the top of this page with your examinee number.
8. 草稿用紙は本冊子から切り離さないこと。
The blank pages are provided for making draft. Do not detach them from this problem booklet.
9. 解答に関係ない記号、符号などを記入した答案は無効とします。
An answer sheet is regarded as invalid if you write marks and/or symbols unrelated to the answer on it.
10. 解答用紙・問題冊子は持ち帰ってはいけません。
Do not take the answer sheets and the problem booklet out of the examination room.

(このページは草稿用紙として使用してよい)
(Blank page for draft)

(このページは草稿用紙として使用してよい)
(Blank page for draft)

問題 1

以下は、長さ m の生物配列 x と長さ n の生物配列 y の局所アラインメントを行うための動的計画法で使われるアラインメントスコアの漸化式である。 x_i は配列 x の i ($=1, \dots, m$) 番目の文字、 y_j は配列 y の j ($=1, \dots, n$) 番目の文字、 $s(a, b)$ は文字 a と文字 b がアラインメントされたときの一致スコアであり、 $d > 0$ とする。以下の問いに答えよ。

$$F(i, j) = \max \begin{cases} 0 \\ F(i-1, j) - d \\ F(i, j-1) - d \\ F(i-1, j-1) + s(x_i, y_j) \end{cases} \quad (\text{ただし } 1 \leq i \leq m, 1 \leq j \leq n)$$

- (1) d はどのような量か、配列の進化に関してどのようなモデルを仮定しているかも含めて簡潔に説明せよ。
- (2) 最大値演算 \max の中の 1 行目の 0 はなぜ必要なのか、この漸化式がどのようなアラインメントを求めるためのものかを含めて説明せよ。
- (3) この漸化式を計算するために、 $i = 0$ もしくは $j = 0$ の場合に対する $F(i, j)$ に初期値を与えることとする。これらの初期値をすべて示せ。
- (4) この漸化式によってアラインメントスコア行列 $F(i, j)$ の値を全て計算する計算複雑度（時間および記憶容量）を示せ。
- (5) この漸化式によって計算されたアラインメントスコア行列 $F(i, j)$ の最大値を持つ要素を探し、トレースバックと呼ばれる手続きを行うことによって、最大のアラインメントスコアを与える局所アラインメントを求めることができる。その場合のトレースバックについて簡潔に説明せよ。

問題 2

生物のゲノム配列は、進化的時間を経るごとにランダムな塩基置換を受ける。ゲノムのある位置での塩基置換を次のような連続時間マルコフモデルでモデル化する。

$$P(b|a, t, \lambda) = [\exp(tS)]_{b,a} \quad S = \lambda \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

ただし $t \geq 0$ は進化時間、自然数 $a, b \in \{1, 2\}$ はそれぞれ 1=プリン (アデニンまたはグアニン)、2=ピリミジン (シトシンまたはチミン) を表し、 $P(b|a, t, \lambda)$ は時刻 0 で塩基 a の時、時刻 t で塩基 b が観測される確率をあらわす。 $\lambda \geq 0$ は塩基置換速度、 $[A]_{b,a}$ は行列 A の b 行 a 列要素を表す。 $\exp(A)$ は行列の指数関数を表し、以下で定義される。但し、 I は単位行列を表わす。

$$\exp(A) = I + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

以下の問いに答えよ。

(1) $M = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ と定義する。 $M^2 = 2M$ を示せ。

(2) c を実数とするとき、次の関係式を示せ。

$$\exp(cM) = I + \frac{e^{2c} - 1}{2} M$$

(3) 関係式 $S = -2\lambda I + \lambda M$ 及び、行列の指数関数に関する次の恒等式

$$\exp(A + B) = \exp(A)\exp(B) \quad \text{if } AB = BA$$

を用い、 $P(b|a, t, \lambda)$ を計算せよ

あるウイルスのゲノム配列を異なる時刻 T_1, T_2 ($T_2 - T_1 = T > 0$) でサンプルした。ここで、 $n_{b,a}$ を塩基 a から塩基 b への変化が観察された配列位置の個数とする ($a = b$ の場合も含む)。また、 C をゲノム配列の全変化 $\{n_{b,a} | a, b \in \{1, 2\}\}$ とする。ここで、時間 T の間に、このような変化 C が起こる確率 $P(C|T, \lambda)$ は、以下で与えられる。

$$P(C|T, \lambda) = \prod_{a,b \in \{1,2\}} P(b|a, T, \lambda)^{n_{b,a}}$$

(4) $L(C, T, \lambda) = \log(P(C|T, \lambda))$ で定義される対数尤度関数 L が最も大きくなるような λ の値を、パラメータ λ の最尤推定値という。停留方程式

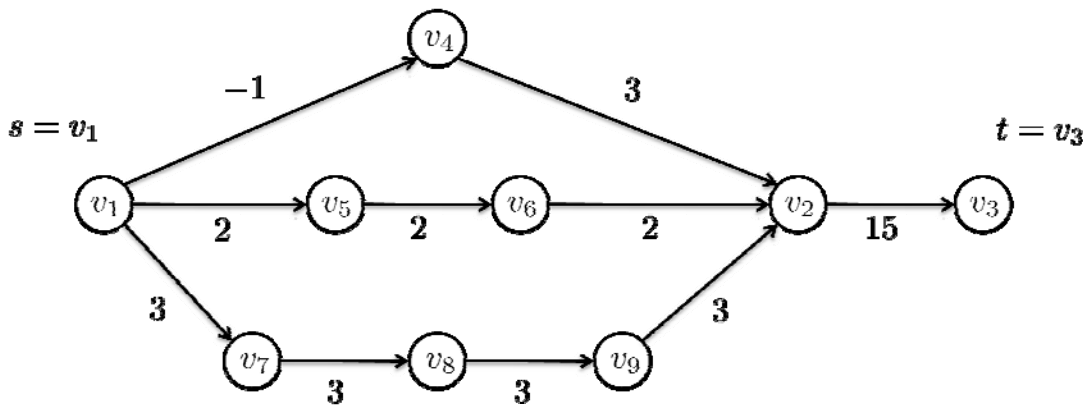
$$\frac{\partial L(C, T, \lambda)}{\partial \lambda} = 0$$

を解き、 λ の最尤推定値とその存在条件を求めよ。

問題 3

頂点集合 V 、辺集合 E 及び辺 $e \in E$ に対する整数の重み関数 $w(e)$ からなる、閉路を含まない重み付き有向グラフ $G=(V, E, w)$ を考える。 G 上のパスの重みとは、パス上の辺に対する重みの和であると定義する。 $s \in V$ 及び $t \in V$ が与えられたとき、 s から t までのパスのうち平均重み（パスの重みをパス上の辺の数で割ったもの）が最小となるようなパスを「平均重み最小パス」と呼ぶこととし、このようなパスを 1 つ求める。

(1) 以下のグラフにおいて、パス $v_1 \rightarrow v_4 \rightarrow v_2 \rightarrow v_3$ の重みを求めよ。



(2) (1) で示したグラフで、 s から t までの平均重み最小パスを求めよ。

(3) G 上において s から t までの平均重み最小パスの重みが定数 C 以下であると仮定する。以下の命題が成り立つことを証明せよ。

「 G' は重み付きグラフ (V, E, w') であり、重み $w'(e) = w(e) - C$ とする。このとき G' において s から t までの重み最小パスの重みが零以下である。」

ただし、重み最小パスとは最小の重みを与えるパスである。

(4) s から t までの重み最小パスを E や V のサイズに対して多項式時間で求めるアルゴリズムを設計せよ。

(5) 平均重み最小パスを $\log(\max_{e \in E} |w(e)|)$ 及び E や V のサイズに対して多項式時間で一つ出力するアルゴリズムを設計せよ。

問題 4

文字列 P において連続した部分文字列が、先頭の文字から始まるとき接頭辞、最後の文字で終わるとき接尾辞と呼ぶ。接頭辞かつ接尾辞となる P より短い部分文字列を「接頭尾辞」と以下では呼ぶ。たとえば $CGCGCG$ において、 CGC は接頭辞、 GCG は接尾辞、 $CGCG$ と CG は接頭尾辞である。長さ 0 の文字列は ε と表記し、 P の接頭尾辞と考える。接頭尾辞を列挙するアルゴリズムを設計するため、以下の問題に答えよ。

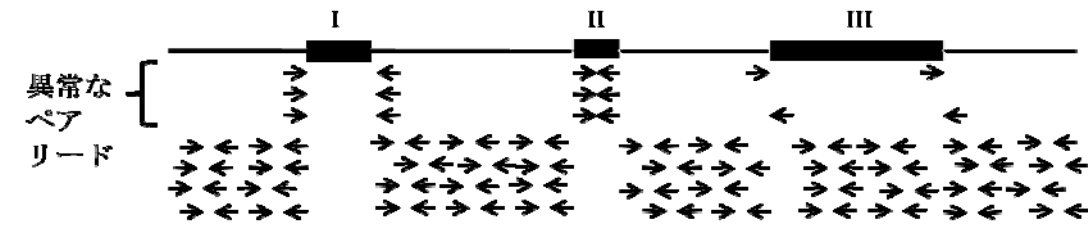
- (1) $P = CGCGCGCG$ と $P = AAAAA$ 各々の場合について接頭尾辞を列挙せよ。
- (2) P の最長の接頭尾辞を $\delta(P)$ と表現する (ただし $\delta(\varepsilon) = \varepsilon$ と約束)。 P に対して関数 δ を $k (\geq 1)$ 回作用させた結果を $\delta^k(P)$ と表記する。 $P = CGCGCGCG$ と $P = CGCGCGCGC$ 各々の場合について、 $\delta(P)$ と $\delta^2(P)$ を求めよ。
- (3) $\delta^k(P)$ ($k \geq 1$) は P の接頭尾辞となることを証明せよ。一方、 P の任意の接頭尾辞は、ある $m (\geq 1)$ により $\delta^m(P)$ と表現できることを証明せよ。
- (4) P のすべての接頭尾辞を、 P の長さに線形比例する時間で列挙するアルゴリズムを設計せよ。

問題 5

下記のゲノム配列（塩基配列は一方の鎖の配列を 5'→3'の方向で記している）はある生物種の全ゲノム配列（reference 配列）の一部である。この配列について以下の問いに答えよ。

GGATCCCCCATATAGGCCCGGAATTCGGAAAAATTTATTTGCTAAGCAGATGGTTGG
AAAATGATTTTCCCAAGGGGTCATGATCGATCATTAAATTATATAATGGACCCCCGTT
 TGACTGGAATTTCCCCGATTACAGATGACTGATGCATGATGGGGCCCTTGACCAAAGG
 CCGCGCGCGCATTTAGGCCCATGGTTTTCCCATGACATGGGGATGAGATATACCCCGGG
 AAGGGATAGTATCGATTAGCAAATTTACCAAGGCCATGACATTGCATCTGAATGCCAAA
 AAATTGGGAGAGACTCTCTCTAGGCCATTGATACCAATTTTGC GTTTTTTAGCAGGATCC

(1) この生物種の個体 X のゲノムシーケンスを全ゲノムショットガンシーケンス法で行い、約 5 kb の鋳型 DNA の両端から約 500 塩基のペアリードを多数取得した。ついで、得られたペアリードを reference 配列にマップしたところ、下図に示したように、予想される 5kb よりもその間隔が長い (I)、短い (II) ペアリード、及び向きが同じペアリード (III) の、異常なペアリードを示す 3 つの領域が存在した。個体 X の I、II、III の領域には reference 配列と比較してどのような構造上の違いがあるか、下記の 4 つの用語 (*) から I、II、III に対してそれぞれひとつ選べ。



→← : ペアリードの間隔が 5 kb → ← : ペアリードの間隔が 5 kb 以上

→← : ペアリードの間隔が 5 kb 以下 → → : ペアリードの向きが同じ

(*) 挿入、欠失、重複、逆位

(2) Reference 配列の下線で示した塩基 G は個体 X では塩基 A であった。このように、同一生物種の個体間で見られる 1 塩基の違いを一般的に何と言うか？

(3) 下線で示した塩基Gを含む236 bpの領域を増幅させるために二重下線の配列を用いてPCRプライマーをデザインした。下記a～fの中で正しいプライマー配列のセットをひとつ選べ。ただし、すべての配列の方向は5'→3'で示してある。

a: AAAATGATTTTCCCAAGGG と CATTGCATCTGAATGCCAAA
b: AAAATGATTTTCCCAAGGG と AAACCGTAAGTCTACGTTAC
c: CCCTTGGGGAAAATCATTTT と CATTGCATCTGAATGCCAAA
d: AAAATGATTTTCCCAAGGG と TTTGGCATTGAGATGCAATG
e: CCCTTGGGGAAAATCATTTT と TTTGGCATTGAGATGCAATG
f: CCCTTGGGGAAAATCATTTT と AAACCGTAAGTCTACGTTAC

(4) 上記のPCRプライマーでPCRを行った。30サイクル後に得られるDNAの分子数は以下のa～fで示した数値の中でどれがもっとも近いか？ただし、1サイクル目に236 bpのDNAが100分子増幅され、2サイクル目以降では増幅したDNA分子がすべて鋳型になると仮定する。なお、 $2^{10} = 1000$ と近似できる。

a: 5×10^7 、b: 5×10^8 、c: 5×10^9 、d: 5×10^{10} 、e: 5×10^{11} 、f: 5×10^{12}

(5) 上記のPCRで得られた236 bpのDNAの重量 (ng) は以下のa～fで示した数値の中で最も近いのはどれか？ただし、DNAの1 bpの分子量 = 600、アボガドロ数 = 6×10^{23} とする。

a: 0.001、b: 0.01、c: 0.1、d: 1、e: 10、f: 100

(6) この生物種のゲノムサイズは約5 MbでmRNAの3'端にはポリA配列がなかった。この生物種と分類上もっとも近縁の生物種を下からひとつ選べ。

線虫、ショウジョウバエ、シロイヌナズナ、メダカ、テトラヒメナ、酵母、麴菌、大腸菌

問題 6

クロマチンの構造変化の研究は近年、急速に進歩している。このクロマチンの構造変化は、遺伝子発現の調節に大きな役割を果たし、種々の生物学的過程において必須である。このクロマチンの構造変化は DNA のメチル化や ヒストン修飾によって影響を受けることが明らかになりつつある。哺乳類において DNA メチル化は細胞分化や X 染色体不活性化、ゲノムインプリンティングに関与している。DNA をメチル化する酵素は、哺乳類では DNA メチル基転移酵素 1 (Dnmt1) と DNA メチル基転移酵素 3 (Dnmt3a, Dnmt3b) の 2 種類が知られており、それぞれは異なる機能をもっている。

下線部に関連した問題について答えよ。

- (1) 哺乳類ゲノムでは DNA メチル化は通常どの塩基で起こるか？また、このメチル化される塩基を高頻度を含むジヌクレオチド配列を 5' 側から書け。
- (2) アセチル化はヒストン修飾の一つのタイプである。他のヒストン修飾のタイプを 3 つ挙げよ。
- (3) DNA メチル基転移酵素 1 と DNA メチル基転移酵素 3 の機能をそれぞれ数行で説明せよ。

問題 7

(1) 以下の用語をすべて用いて、2種類の細胞骨格について説明せよ。同じ用語を何回使っても構わない。

ミオシン - 微小管 - キネシン - 中心体 - アクチン - ダイニン - マイクロフィラメント - 鞭毛 - チューブリン - 筋肉 - プラス端 - マイナス端

(2) タキソールは、微小管に強く結びついて安定化させ、微小管の重合形成を促進する。一方コルヒチンは、微小管の形成を阻害する。正反対の性質を持つ2つの薬剤は、どちらも抗ガン剤として用いられる。これら正反対の薬剤が、どうしてガン細胞に対して同じように強い毒性を発揮するのかを数行で説明せよ。

(3) 2種類の細胞内シグナル伝達系について、以下の用語(1～15)を、A: Gタンパク結合受容体を介したシグナル伝達、B: 受容体チロシンキナーゼを介したシグナル伝達の、どちらにより深く関与しているかを記せ。番号と答えを20-A, 21-Bのように書け。

1. アダプタータンパク - 2. 環状 AMP - 3. Ca^{2+} - 4. MAP キナーゼ -
5. 小胞体 - 6. 7回膜貫通受容体タンパク - 7. RAS -
8. アデニル酸シクラーゼ - 9. 低分子量 G タンパク(低分子量 G T P 結合タンパク) -
10. 三量体型 G タンパク - 11. α サブユニット - 12. ホスホリパーゼ C -
13. イノシトール 3 リン酸 (IP3) - 14. プロテインキナーゼ A -
15. セカンドメッセンジャー

問題 8

(1) Aで挙げたそれぞれのタンパク質に最も関係の深い分子をBから選べ。

A. ヘモグロビン, カルボキシペプチダーゼ, リゾチーム, ロドプシン

B. レチナール, ヘム, 亜鉛イオン, 多糖

(2) Cで挙げたそれぞれのタンパク質の構造的特徴として最も適当なものをDから選べ

C. ノイラミニダーゼ, コラーゲン, CAPタンパク質, α -ケラチン

D. 2量体, 3本鎖らせん, より合わせコイル, 4量体

(3) 下記の用語をすべて用いて酵素のアロステリック性を説明せよ。

調節分子 活性部位 コンホーメーション変化 基質

(このページは草稿用紙として使用してよい)
(Blank page for draft)

(このページは草稿用紙として使用してよい)
(Blank page for draft)