

دانشگاه صنعتی شریف – دانشکده مهندسی کامپیوتر

آزمون پایان ترم درس مقدمه‌ای بر بیوانفورماتیک

وقت: ۲,۵ ساعت

۱۳۹۸/۴/۳

۱. به هر سوال زیر در حد دو-سه سطر پاسخ دهید:

الف) مفهوم p-value چیست؟ اگر p-value یک آزمون فرض 0.001 باشد چه نتیجه‌ای می‌گیرید؟

ب) در یک مطالعه‌ی دارویی ۸ آقا و ۱۰ خانم بیمار شرکت کرده‌باشند و از کل افراد، ۱۲ نفر بهبود یافته‌اند. اگر اثر دارو ربطی به جنسیت بیمار نداشته باشد، چقدر احتمال دارد که از افراد بهبود یافته دقیقاً ۱ نفر آقا باشد؟

پ) می‌خواهیم اثر استفاده از روغن حلزون بر مدت زمان بهبود زخم را مطالعه کنیم. برای این کار، ۱۰۰ نفر شرکت‌کننده که زخم داشته‌اند را به دو گروه ۵۰ نفره تقسیم کرده‌ایم، از شرکت‌کنندگان گروه اول خواسته‌ایم روی زخم خود روغن حلزون بمالند و از گروه دیگر خواسته‌ایم این کار را نکنند. مدت زمان بهبود زخم را برای هر بیمار یادداشت کرده‌ایم. عددهای به‌دست آمده از یک توزیع شناخته‌شده‌ی آماری پیروی نمی‌کند. چه روشی را پیشنهاد می‌کنید برای آن که بتوانیم بررسی کنیم که روغن حلزون روی بهبود زخم اثر معنی‌دار دارد یا خیر؟ این روش را چگونه باید انجام دهیم و نتایج آن را چگونه استنباط کنیم؟

ت) در قسمت قبل، اگر بدانیم اعداد به‌دست آمده در هر گروه از توزیع نرمال و با واریانس‌های نسبتاً مشابه پیروی می‌کنند آیا روش بهتری را می‌توانید پیشنهاد کنید؟

ث) ساختارهای اول، دوم و سوم پروتئین را توضیح دهید.

ج) اگر هنگام ترجمه RNA به پروتئین شیفت خوردن یک حرف بود (آمینو اسید اول سه از حرف اول RNA و آمینو اسید دوم از حرف‌های ۲ تا ۴ و آمینو اسید سوم از حرف‌های ۳ تا ۵ ... ترجمه می‌شد) چه مشکلی رخ می‌داد؟

چ) مساله‌های دور هامیلتونی و تور اویلری را توضیح دهید. زمان اجرای الگوریتم بهینه برای هر یک را ارائه دهید.

ح) در هنگام اسمبلی ژنوم، چگونه مساله دور هامیلتونی را به تور اویلری تبدیل می‌کنیم؟

۲. یک HMM با شرایط زیر فرض کنید:

$$\Sigma = \{1,2,3\}$$

$$Q = \{A, B, C\}$$

$$\pi_A = 1, \pi_B = 0 \text{ and } \pi_C = 0$$

مقادیر transition probabilities و emission probabilities از جدول زیر بدست می آید:

	A	B	C	1	2	3
A	0.25	0.25	0.5	0.5	0.25	0.25
B	0.5	0.00	0.5	0.5	0.0	0.5
C	0.33	0.33	0.33	0.0	0.5	0.5

الف) نمودار حالت HMM فوق را رسم کرده و احتمال گذر (transition probabilities) را روی هر یال نشان دهید.

ب) همه مسیرهای حالت ممکن را برای مشاهده  $O = 1,3,1$  بدست آورید.

ج) مقدار  $P(O|H)$  را محاسبه کنید.

د) محتملترین مسیر برای مشاهده  $O = 1,3,1$  و مقدار احتمال آن را بدست آورید.

۳. فرض کنید برای یافتن motif رشته های DNA داده شده به صورت زیر align میشوند.

```

123456789
12345... CGGGGCTATcCAgCTGGGTCGTCACATTCCTTT...
TTTGAGGGTGCCCAATAAagGCAACTCCAAAGCGGACAAA
GGATGgAtcCTGATGCCGTTTGACGACCTA...
AAGGAaGCAACcCCAGGAGCGCCTTTGCTGG...
AATTTTCTAAAAAGATTATAATGTCGGTCCtTGgAACTTC
CTGCTGTACAACCTGAGATCATGCTGCATGcCAtTTTCAAC
ACATGATCTTTTGATGgcACTTGGATGAGGGAATGATGC

```

الف) آرایه مکان شروع Motif (starting position array) را بدست آورید.

ب) ماتریس profile را کامل کنید.

ج) رشته consensus را بدست آورید.

Alignment Matrix:

A	T	C	C	A	G	C	T
G	G	G	C	A	A	C	T
A	T	G	G	A	T	C	T
A	A	G	C	A	A	C	C
T	T	G	G	A	A	C	T
A	T	G	C	C	A	T	T
A	T	G	G	C	A	C	T

Profile Matrix:

A							
T							
G							
C							

Consensus String: \_\_\_\_\_

د) اگر بدانیم  $\mu = -1$ ,  $\delta = -2$  (Mismatch penalty  $\mu$  و Indel penalty  $\delta$ ) و  $\text{match score} = 1$  ، حداقل مقدار alignment score با فرض رشته مرجع consensus ، به ازای کدام رشته DNA بالا بدست می آید.

۴. با استفاده از دو روش UPGMA Tree و Neighbor-Joining Tree ، درخت فیلوژنی چهار گونه باکتری زیر را استخراج کردیم.

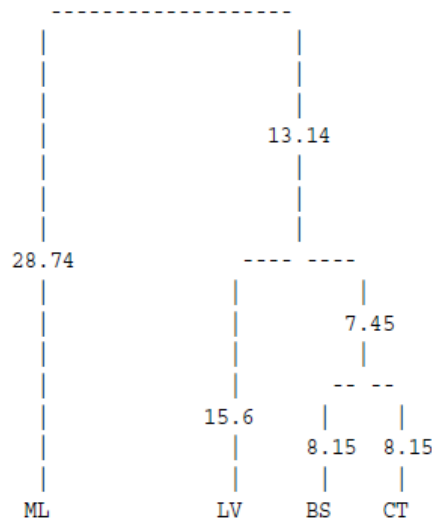
M10815 *Bacillus subtilis* (BS)

X02713 *Lactobacillus viridescens* (LV)

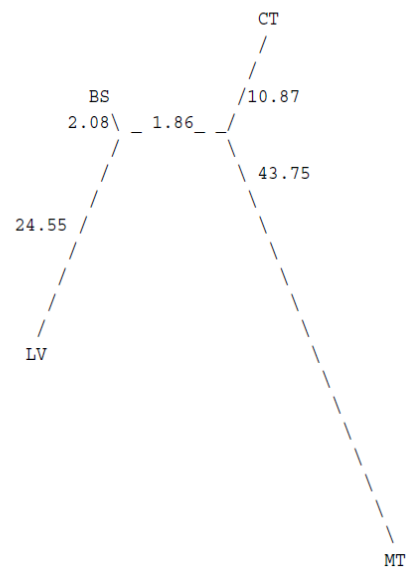
M58416 *Clostridium tyrobutyricum* (CT)

K02683 *Micrococcus luteus* (ML)

**UPGMA Tree:**



**Neighbor-Joining tree:**



الف) آیا دو درخت فوق با هم معادلند، با استدلال توضیح دهید.

ب) با فرض ماتریس فاصله زیر (distance matrix) ، فاصله *Lactobacillus viridescens* و *Micrococcus luteus* را بدست آورید.

1 BSubtilis

2 CTyro

3 LViridescens

4 MLuteus

	1	2	3	4
1	0.00	16.31	26.63	46.18
2		0.00	35.77	54.62
3			0.00	71.66
4				0.00

ج) فاصله دو گونه فوق را روی هر یک از دو درخت UPGMA و Neighbor-Joining بدست آورده و با مقدار بدست آمده از قسمت (ب) مقایسه کنید. به کدام نزدیکتر است؟

۵. مراحل یافتن محل یک خوانده‌ی ACGA را در یک ژنوم AACTTAACGACCTGTA با استفاده از داده‌ساختار BWT شبیه‌سازی کنید و به طور کامل توضیح دهید.

الف) از ژنوم چطور BWT می‌سازیم؟

ب) ژنوم را چطور به داده‌ساختاری برای جستجو تبدیل می‌کنیم؟ آن داده‌ساختار را برای ژنوم داده شده بنویسید.

ج) خوانده را چطور در داده‌ساختار به دست آمده جستجو می‌کنیم؟