

# 國立陽明大學 生化所九十學年度 生物資訊學 期中考

姓名:

系所別 (學程):

學號

請一律於本答案紙作答，答案紙用完為止，請節省使用，最好不要寫至背面：

1	Identity 71.6%	Similarity 85%	2 B	3 D	4 94	5 B
6 A			7 A	8 A	9 B,C	10 C
11 C			12 D	13 B	14 B	15 A
16 D			17 B	18 B	19 A	20 B

以下為問答題 (勿長篇大論)

1.

A transcription factor may regulate the expression of several genes by binding to their promotor regions. Variations in the sequence composition of the binding sites may determine the specificity of transcription regulation. Using consensus sequence to describe the binding site may not be proper because it may lose those subtle features. Sometimes the variations among sequences from different genes may be very large. Matrix calculation may preserve all features by estimating the site-specific probability of each nucleotide and thus it is suitable for searching transcription factor binding sites.

2

Though profile has considered the information of site-specific probability of each amino acid, it does not tolerate the existence of a gap. In addition to considering the site-specific probability of amino acids, Profile HMM can tolerate the existence of gaps.

3.

A. 103355-103913 → 1-559

B. 118521-118607 → 558-644

C. 105274-105436 → 638-800

D. 105967-106139 → 798-961

E. 106550-106772 → 1087-1309

F. 108121-108497 → 1333-1709

According to the re-arrangement of the mRNA fragments, obviously the genomic sequences B and C are in the wrong order and they must be re-ordered. Consequently, the genomic sequence is belonged to phase I.

國立陽明大學 生化所九十學年度 生物資訊學 期中考

姓名: \_\_\_\_\_ 系所別 (學程): \_\_\_\_\_ 學號 \_\_\_\_\_

1. 下列排比的%identity及%similarity為何?

```
Query: 4 MYAYIREAWKRPYEGYVVGELMWHRLQKWRREPAVVRI PRPTRLDRARALGYKAKKGIIVV 63
MY YIREAWK P + YVG+L+ R+ KWRREPAVVRI RPTLDRARALGY+AK+G ++V
Sbjct: 3 MYKYIREAWKSPKKS YVGQLLKQRMKWRREPAVVRI ERPTLDRARALGYQAKQGYVIV 62
```

2. THC 與 UniGene 的最大差別在於 1) THC 是 EST 以 3'-end 為中心，來做分群的依據，所得到的結果；2) THC 是將 EST 組合出的 contig；3) THC 是由 UniGene 所衍生的加值資料庫；4) THC 更新的速度快於 UniGene
3. 下列有關 EST 的？速何者為真？1) Merck project 所做出的 EST 是以鐳射切割出的組織，做成 library 後，取得的部份序列；2) CGAP 所產生的 cDNA 殖株近似於 full length cDNA；3) 目前 UniGene 是 IMAGE 下的重要計畫；4) IMAGE 的殖株中約 25% 是有問題的
4. 下列的 profile 應得多少分

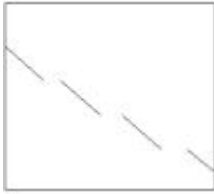
	C	T	A	T	A	A	T	C
A		-38	23	1	12	8	-48	
C		-15	-38	-8	-12	-3	-32	
G		-13	-48	-13	-5	-8	-48	
T		17	-36	15	-9	-6	19	

5. 若欲以 EST 序列比對蛋白質資料庫，應使用 1) blastp；2) blastx；3) tblastx；4) tblastn；5) blastn
6. BLAST 的輸出檔中，E 值代表 1) the probability to find a sequence that is similar to the query by chance；2) extra information related to this sequence；3) entropy of the information content in the sequence alignment；4) overflow when performing the sequence alignment
7. 如果你想要增加 fasta 或 blast 的執行速度，你可以 1) 增加 word size；2) 減少 word size；3) 增加 E-value cutoff；4) 減少 E-value cutoff
8. 如果你想要偵測微弱的(weak)同源性 (homology)，下列程式何者可以提供較佳的結果: 1) Smith-Waterman；2) FastA；3) Blast；4) rps-blast

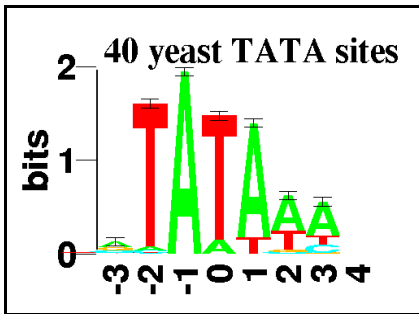
國立陽明大學 生化所九十學年度 生物資訊學 期中考

姓名: \_\_\_\_\_ 系所別 (學程): \_\_\_\_\_ 學號 \_\_\_\_\_

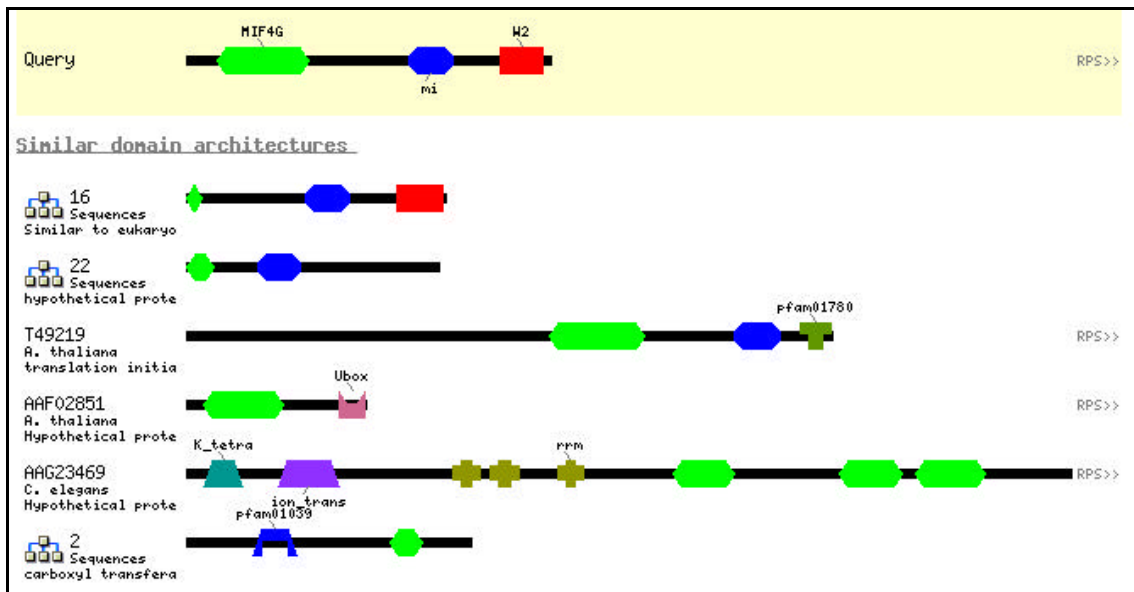
9. 下圖為 cDNA 與 genomic sequence alignment，請問下列？述何者為錯誤 1) 縱軸為 cDNA；2) 縱軸為 genomic sequence；3) intron 的長度為 exon 的 10 倍；4) 此圖之有 4 個 exons



10. 下圖是 1) consensus sequence; 2) sequence profile; 3) sequence logo; 4) HMM profile



11. 下圖是 DART 程式的部分輸出檔，下列何種？述為真？ DART 會 1) 列出具相似序列的蛋白質；2) 列出具相似模組的蛋白質；3) 列出具相似模組組成的蛋白質；4) 只列出一序列之所有蛋白質模組

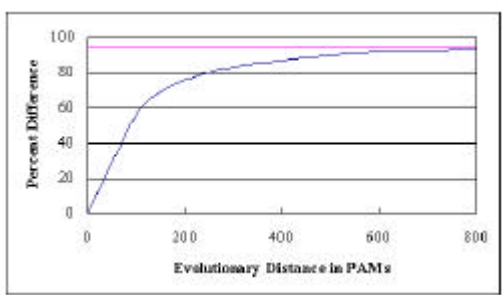


12. 有關 PAM matrix 的？述何者為假？ 1) PAM=accepted point mutation；2) 只有 PAM1 是統計真實序列之數據；3)  $PAM_{250} = (PAM1)^{250}$ ；4)  $PAM_{250} = 250 \times (PAM1)$

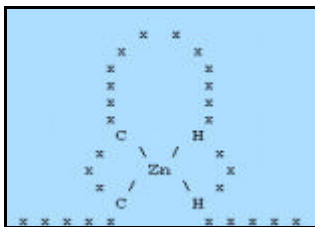
國立陽明大學 生化所九十學年度 生物資訊學 期中考

姓名: \_\_\_\_\_ 系所別 (學程): \_\_\_\_\_ 學號 \_\_\_\_\_

13. 參考下圖，若欲比較在序列上有 50%相似的序列，以下列何者較為適合? 1) PAM30 ; 2)PAM80 ; 3)PAM150 ; 4)PAM250



14. 找不到序列相似性時，可以靠哪一種方式來猜測基因是屬於同一反應路徑的? 1) motif analysis; 2) microarray and clustering analysis; 3) protein-protein interaction; 4) combination analysis
15. Blosum matrix 比 PAM matrix 好的原因是 1) Blosum 是收集具有差異，而非高度相似的序列所做出的 matrix ; 2) Blosum1 的誤差比 PAM1 小 ; 3) Blosum45 相對於 PAM250，因此誤差放大的比例較小 ; 4) PAM matrix 是使用舊的資料庫算出的 matrix
16. 下圖為 Zn-Finger 的一個示意圖，請問下列何者為 Zn Finger 的 pattern? 1) C-2-C-2-[LIVMFYWC]-8-H-2-H ; 2) C-x(2,4)-C-[LIVMFYWC]-H-x(3,5)-H ; 3) C-x(2,4)-C-x(14,16)-H-x(4,6)-H ; 4) C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H



17. 使用 blast 尋找相且近於同的短序列時，將 E 變大為 1000 且將 word size 變小為 7，其目的為 1) 增加計算速度 ; 2) 避免找不到介於 7~11 的短序列 ; 3) 減少收集不太相似的序列的機會 ; 4) 減少因隨機產生的相似的機會
18. 用 pairwise alignment 對 cDNA 與 genomic DNA 進行比對時，若欲尋找基因剪接點，下列何種 gap penalty 之設定最可能有好的結果 (E 表示 gap extension penalty, I 表示 gap insertion penalty) 1) I 大 E 大 ; 2) I 大 E 小 ; 3) I 小 E 大 ; 4) I 小 E 小

