

## 演習問題

1. 以下の図は、シロイヌナズナのジベレリン受容体 **GID1A** の配列 (345 残基) をクエリとして NCBI の **blastp** で **Swiss-Prot** データベースに対して **BLAST** 検索を行った結果、ヒットしたシロイヌナズナのカルボキシエステラーゼ **CXE2** と推定される配列 (314 残基) について、検索結果のサマリーと配列アラインメントを示したものである。アミノ酸置換マトリックスは **BLOSUM 62**、**gap existence penalty = 11**, **gap extension penalty = 1** を用いた。これについて、後の問いに答えよ。

	Score	Expect	Method	Identities	Positives	Gaps
	133 bits(334)	4e-35	Compositional matrix adjust.	86/292(29%)	151/292(51%)	26/292(8%)
	①	②		③	④	⑤
Query	58	NPVDGVVFSFDVLIDRRINLLSRVYRPAYADQEQPPSILDLEKPVGDIVPVLFFHGGSF				117
		+P+ GVFS D++I+ + L +R+YRP QP G +P++L+FHGG+F				
Sbjct	38	DPITGVVFSKDIIIIEPKTGLSARIYRPFPSI---QP-----GQKIPLMLYFHGGAF				83
Query	118	AHSSANSAIYDTLCRRVLGKCKCVVSVNYRRAPENPYPCAYDDGWIALNWNVS--RSWL				175
		SS + Y T ++V + VSVNYR APE+P P AY+D W AL + + W+				
Sbjct	84	LISSTSFPSYHTSLNKIVNQANVIAVSVNYRLAPEHPLPTAYEDSWTALKNIQAINPWI				143
Query	176	KSKKDSKVHIFLAGDSSGGNIAHNVALRAGES--GIDVLGNILLNPMFGGNERTESEKSL				233
		D +FL GDS+G NI+H++A RA +S + + G +++P F G + +E				
Sbjct	144	NDYADLD-SLFLVGI(S)GANISHHLLAFRAKQSDQTLKIKIGIGMIHPYFWGTQPIGAEIKD				202
Query	234	DGKYFVTVRDRDWDYWKAFLEPEGEDREHPACNPFSPRGKSLEGVSFPKSLVAVGLDLIRD				293
		+ + + D+W+ P + + P NPF+ L G+ + ++ VA D++ +				
Sbjct	203	EARKQMV----DGWWEFVCPSEKGSDDPWINPFADGSPDLGGLGCERVMITVAE(D)ILNE				258
Query	294	WQLAYAEGLKKA--GQEVKLMHLEKATVGFYLL-PNNNHFNVMDEISAFVN 342				
		Y E L K+ +V++M ++ E++ P+ + ++ ++ F+N				
Sbjct	259	RGKMYERLVKSEWKGKVEIMETKEKDHVPHIFEPDCDEAMEMVRCLALFIN 310				

- (1) 以下の(a)~(g)を埋めよ。(a), (f) は{ }内のカンマで区切った選択肢のうち1つを選べ。

BLAST を実行した結果、シロイヌナズナの **GID1A** の (a) {全長, N 末側を除いた部分, C 末側を除いた部分, N 末・C 末側を除いた部分} の (b) 残基の配列と、シロイヌナズナの **CXE2** の (c) 残基の配列のアラインメントが得られた。アラインメントの長さはギャップを含めて (d) 残基で、配列一致度は(e)である。E-value は (f) { $2 \times e^{-34}$ ,  $2 \times 10^{-34}$ } (e はネイピア数、自然対数の底、(f)の解答は e か 10 で答えること) で、これは (g) 以上のビットスコアが得られる期待値を示している。

- (2) 上の図の①~⑤の数値のうち、クエリ配列の長さに依存しないものをすべて挙げよ (Google Forms で該当する番号をすべて選択せよ)。
- (3) 上の図で、クエリ配列 (Query) と検索されたデータベースの配列 (Sbjct) の間には、2つの配列のアミノ酸の対応関係が記載されている。そこに記載されている「+」は何を意味しているか、具体的に説明せよ。
- (4) カルボキシエステラーゼ **CXE2** の触媒 3 残基は、図中の赤丸で示した残基番号 158 のセリン、残基番号 254 のアスパラギン酸、そしてヒスチジンである。このヒスチジンがジベレリン受容体ではバリンに置換されたことで触媒能を失い、受容体タンパク質として機能するように進化したと考えられている。この **CXE2** のヒスチジンの残基番号を答えよ。

(5) 以下は、本問の BLAST 検索でヒットした配列のリストである。

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident
1	<a href="#">RecName: Full=Gibberellin receptor GID1A; AltName: Full=AtCXE10; AltName: Full=Carboxylesterase</a>	711	711	100%	0.0	100.00%
2		615	615	100%	0.0	85.51%
3		513	513	99%	0.0	69.19%
4		450	450	99%	2e-158	61.82%
5		219	219	99%	7e-68	36.57%
6		157	157	81%	7e-44	34.60%
7		145	145	84%	1e-39	31.94%
8		139	139	80%	2e-37	33.33%
9		<a href="#">RecName: Full=Probable carboxylesterase 2; AltName: Full=AtCXE2 [Arabidopsis thaliana]</a>	133	133	82%	4e-35
10		133	133	71%	6e-35	35.34%
11		131	131	85%	2e-34	32.40%
12		130	130	63%	4e-33	32.95%
13		127	127	88%	8e-33	29.10%
14		125	125	84%	3e-32	31.61%
15		127	127	78%	6e-32	33.21%
16		125	125	73%	9e-32	32.83%
17		124	124	88%	1e-31	31.03%
18		123	123	76%	2e-31	32.86%
19		124	124	81%	2e-31	30.82%
20		122	122	69%	3e-31	34.51%

左端に、各配列の番号を記す。配列 1 はクエリ配列、配列 9 はアラインメントを示したシロイヌナズナの CXE2 である。この検索で、シロイヌナズナの他の複数のジベレリン受容体とイネのジベレリン受容体 GID1 が検索でヒットした。これらはカルボキシルエステラーゼとは（類縁ではあるが）別なファミリーに属する。また、シロイヌナズナのジベレリン受容体どうしは、イネのジベレリン受容体よりも高い配列一致度を示した。イネの GID1 は、1~20 のどの配列に相当するか、番号で答えよ。

- (6) 検索対象のデータベースを「Non-redundant protein sequences」(NCBI のゲノム配列の翻訳によって収集した非冗長データベース) としたとき、Swiss-Prot データベースの場合に比べ、(同じ E-value の基準で打ち切ったときの) ヒット数は増えるか減るかどちらが予想されるか。また、そう答える理由を説明せよ。

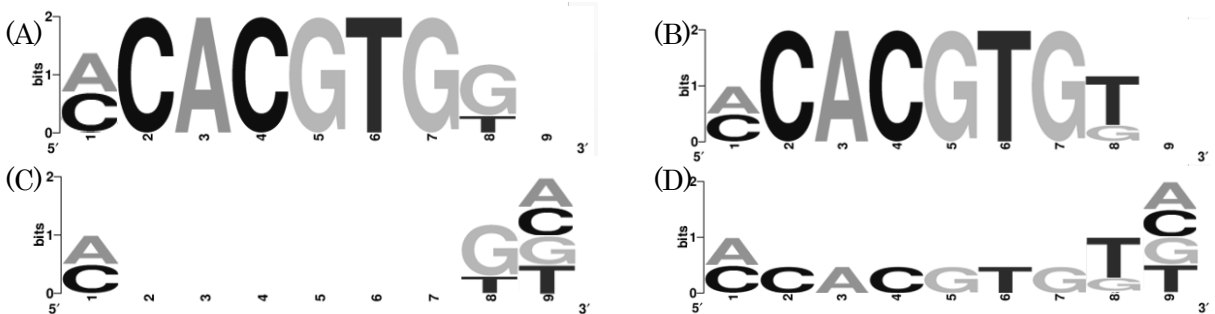
2. 以下の図は、4つの植物のアブシシン酸応答配列 (ABARE) のマルチプルアラインメントを示したものである。後の問いに答えよ。

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
AtNCED3	C	C	A	C	G	T	G	T	T
VvNCED1	A	C	A	C	G	T	G	G	A
AhNCED1	C	C	A	C	G	T	G	T	G
OsNCED3	A	C	A	C	G	T	G	T	C

- (1) 塩基が一致、不一致のスコアをそれぞれ+1、-1として、これらの配列のペアワイズアラインメントのアラインメントスコアを求め、次の表の(a)~(e)を埋めよ (配列 AtNCED3 と配列 VvNCED のアラインメントスコアはすでに記入している)。

	AtNCED3	VvNCED1	AhNCED1	OsNCED3
AtNCED3	—	3	(a)	(b)
VvNCED1	—	—	(c)	(d)
AhNCED1	—	—	—	(e)

- (2) マルチプルアラインメントを、ペアワイズアラインメントを段階的に適用して構築するプログレッシブアラインメントの手法の例を1つ挙げ、その手順を説明せよ。
- (3) (2)の手法で、4つの配列のうち、最初にペアワイズアラインメントが形成される2つの配列を示せ。
- (4) 配列位置1~9のうち、情報エントロピーが最大の配列位置と、その位置の情報エントロピーの値を求めよ。  $0 \log_2 0 = 0$  とする。
- (5) このマルチプルアラインメントのシーケンスロゴとして適当なものを以下の(A)~(D)から選べ。



## 解答

1.

(1) (a) N, C 末側を除いた部分 (b) 285 (c) 273 (d) 292 (e) 29% (f) 10 (g) 133

(2) ③, ④, ⑤

(3) アミノ酸は異なるが、置換マトリックスにおいてスコアが正のもの

(4) 286

(5) 4

(6) ヒット数が増える。なぜなら、Swiss-Prot を含む、より大きなデータベースだから。

2.

(1) (a)7 (b)5 (c)3 (d)5 (e)5

(2) ツリーアラインメント

1.すべての配列ペアの組み合わせに対してペアワイズアラインメントを行い、配列間の類似度スコアまたは配列間の距離を計算する。

2.そのうちペアワイズアラインメントのスコアが最大のものでペアを作り、アラインメントを構築する。

3.構築したアラインメントを配列に加え、新たな全ての配列ペアのスコアを再計算する。

4.これを繰り返す。

これを全ての配列のアラインメントが得られるまで繰り返す。

これにより案内木を作成し、アラインメントを構築する。

(3) AtNCED3 と AhNCED1

(4) 配列位置 9, 情報エントロピー 2

(5) (B)