

問 1. 以下の下線部を埋めよ、

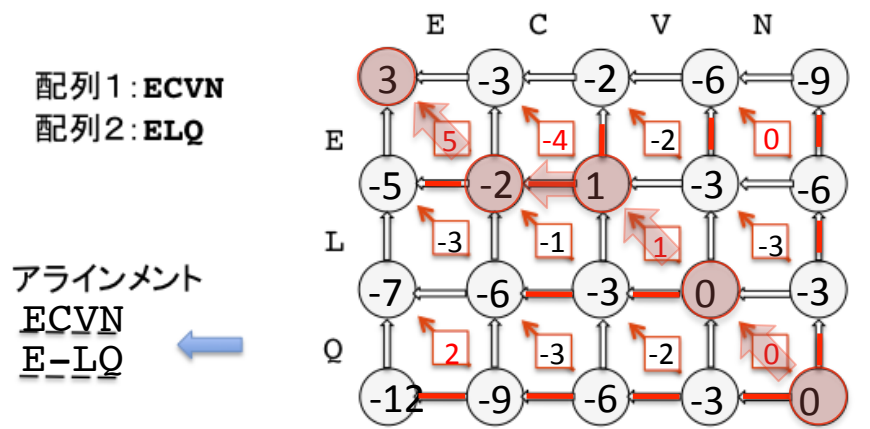
- (1) タンパク質は 20 種類のアミノ酸が鎖状に繋がって出来ている。ひとつひとつのアミノ酸をアルファベットで表記することで、アミノ酸配列を表すことができる。こうしたアミノ酸配列を、由来する生物種、タンパク質としての機能などの付加情報とともに記録したデータベースとしては、例えば Swissprot/Uiprot などがある。機能の分からないタンパク質のアミノ酸配列があるとき、このデータベースに対してプログラム BLAST を用いて、ホモロジー検索を行うことで類似したタンパク質を探し、機能推定を行うことができる。
- (2) (1) で行った検索によりえられたアミノ酸配列がもとのクエリー配列と相同性 (ホモロジー) があるかどうかを判定するためには、ペアワイズアラインメントで一致するアミノ酸残基の割合を示す 配列相同性 という値、あるいは e-value という値を調べる。目安として、前者が、20% 以上、あるいは後者が 0.0001 以下であれば、日本の配列の間に相同性 (ホモロジー) があるとみなしてよい。
- (3) 二つのタンパク質が共通祖先に由来するために類似している場合、一方がもう一方の ホモログ であるという。そのうち、異なる生物にあって同じ機能を有するものを オーソログ、同じ生物でわずかに異なる機能を有するものを パラログ と呼ぶ。
- (4) 原核生物 (バクテリア) のゲノム配列からタンパク質をコードする領域を見つけるためのひとつの方法として、十分な長さにわたって、ストップコドン の出現しない、いわゆる ORF を探す方法がある。
- (5) 相同性のある複数のタンパク質アミノ酸配列をマルチプルアラインメントしたときに、配列間で同じアミノ酸がよく揃っている部分を 保存領域 と呼ぶ。この領域は タンパク質の機能・構造の維持 に重要であると考えられる。

問 2. 右下に示すBLOSUM62スコアマトリクスを用いて、以下のアラインメントについてスコアを計算せよ。
ただし、ギャップ挿入ペナルティを -11 ギャップ延長ペナルティを -1 とする。計算の過程も記すこと

- 1) **YVAUGHNCQECC** スコア = 37
YIAWHDC EA E C C
7+3+4+11-2-3-4-1+5+9+9
- 2) **YVAUGHNCQECC** スコア = 38
YIAW-HDC EA E C C
7+3+4+11-11+8+1+9+2-1-4+9
- 3) **YVAUGHNC-QECC** スコア = 43
YIAW-HDC EA E C C
7+3+4+11-11+8+1+9-11-1+5+9+9

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-2	-1	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

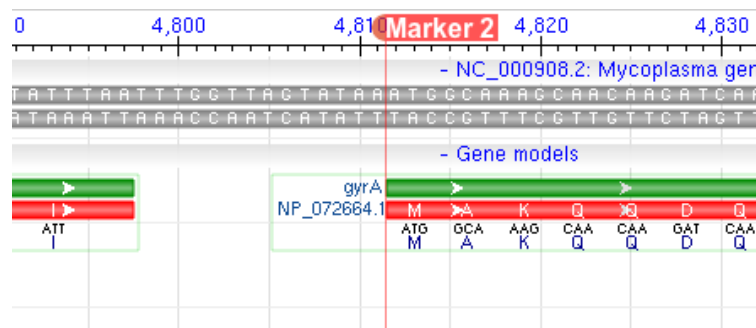
問 3. 動的計画法を使い、右に示した図を利用して配列1と配列2のペアワイズアラインメントをせよ、ただし、上記のBLOSUM62スコアマトリクスを用い、ギャップ挿入ペナルティ、ギャップ延長ペナルティをともに -3 とする。



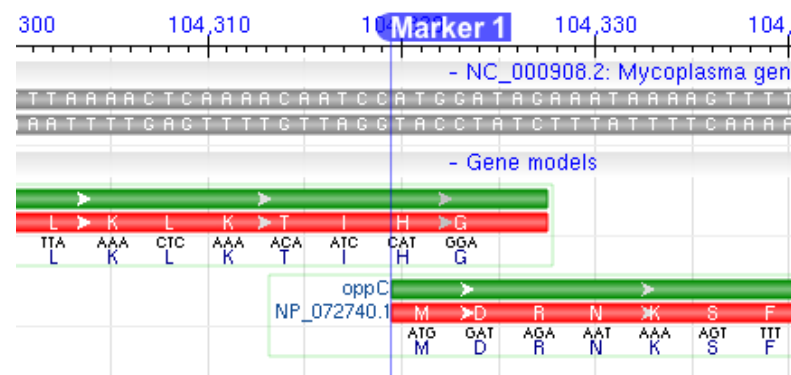
問 4. 以下のそれぞれのPROSITE配列シグネチャを下のアミノ酸配列中に探し、その位置を示せ

- 1) G - R - x - N - [LIV] - I - G - [DE] - H - x - D - Y
VPRGSHMSIVVENSTVLSALTEKFAEVFGDTKEVEYFFS **PGRINLIGEHTDY** NGGYVFPASITIGT
- 2) D - P - x - F - [LIVMFYW] - x(2) - H - x(3) - D
HNWGHVMMANITDPDHRFQENPGVMSDTSTSLR**DP**IFYRWHRFIDNIFQEHKKSFPYTKHEELSFPGVEVVG
- 3) [SAPG] - [LIVMST] - [CS] - [STACG] - P - [STA] - R - x(2) - [LIVMFW](2) - [TAR] - G
LAAGGLRFTDFYVPV **SLSTPSRAALLTG**RLPVRMGMPGVLPSSRGGLPLEEVTVAEVLAAARGYLTGMAGKWHL

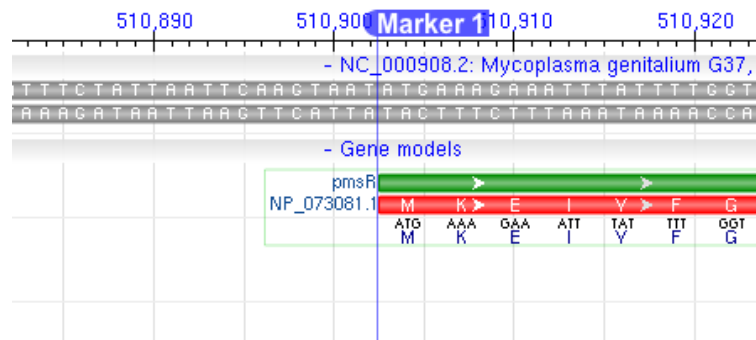
問 5. 以下に示したNCBI gene viewerの表示を見て、例にならってそれぞれの遺伝子の開始点前後6塩基ずつと開始点以後のアミノ酸配列3残基を記せ
相補鎖配列についてもFlip表示をせずに順方向で表示されていることに注意すること、



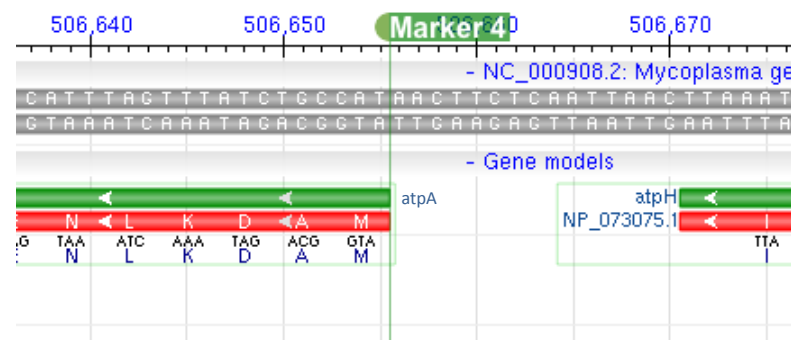
例)gyrA 塩基配列: GTATAA-ATGGCA アミノ酸配列: MAK



1)oppC 塩基配列: CAATCC-ATGGAT アミノ酸配列: MDR



2)pmsR 塩基配列: AGTAAT-ATGAAA アミノ酸配列: MKE



3)atpA 塩基配列: GAAGTT-ATGGCA アミノ酸配列: MAD

問 6. 以下に示したのはGenBankフォーマットによる、Mycoplasma Genitaliumゲノム中のある遺伝子(MG_011)の情報と関係する塩基配列領域を抜粋したものである。これを見て、MG_011のコードするアミノ酸配列の初めの5残基と、塩基配列の初めの9塩基を記せ。

```

コード領域情報
CDS             complement(12701..13564)
                /locus_tag="MG_011"
                /note="identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=4
                /product="hypothetical protein"
                /protein_id="NP_072671.1"
                /db_xref="GI:12044861"
                /db_xref="GeneID:875236"
                /translation="MGKIKKNRKALVVYDNKDDFEKNQTFALSILKELQKKLNAEV
                LLENKDINFEAKINEAELILNRSRKVDFLKTNNQINTFLVNPFFNVVFIANDKYETK
                WLKQNRFLTIVNSLLSKETIKSFPVIVKRNHSHGGKDVHLVNSADEIKHLNINATEW
                IVQPFISIGTVEYRAYILFGKIIKVIKKISNANQFKANFSQGAEVSLFKLWFTKRKI
                KKIARLRREGYYAIDFFLNRYNRYVINEIEDAAGARALVQLCPDLNITKIIIRTIISK
                FKKFLKKKLIS"
    
```

```

塩基配列(一部を抜粋)
12661 tgatttgttt ttattaaaca actcaaattg atcagcggtt ttaactaatc aacttctttt
12721 ttaagaattt tttaaattta ctaataattg ttctgataat tatttttagtg atattttaat
12781 ctggacaaag ctgaactaaa gctctcgac cagcagcatt ttcaatttca ttaacaataa
12841 ccctattata tctattttaa aagaagtcaa tagcataata accttccctt aggcgtttag
12901 ctattttctt tatttttctt ttagtaaatc actttaattt aaacaaggaa acttcagcac
12961 cttgtgaaa gttagcttta aattgattag cattagaaat ttttttaata actttaatta
13021 tttttccaaa caaaatataa gcacgatatt caactgtgcc aattgataaa aaaggttgaa
13081 caattcattc tgttgcattt tcaatgttta aatgtttgat ctgcgtagca ctattaacta
13141 aatgtacatc ttttccaccg tgtgaattac gtttcttaac gatgacagga aatgatttga
13201 ttgtttcttt actaagaaga gaagaattga cagttagaaa tctattttgt tttaatcatt
13261 tataatgttc gtatttatcg ttgtctataa aaacaacatt aaaaggatta actaaaaaag
13321 tatttatttg attatttggt tttaaaaaat ctacttttct tgaacgattt aaaatcaatt
13381 cagcttcatt aatttttagc tcgaaattaa tgtctttatt ttcaagtaat aagacttcag
13441 catttagttt tttcttttgg aattccttga ttgacttaa agcaaatggt tgattttttt
13501 caaaatcatc cttgttgca taaacaacta atgcttttct gttttttaat ttaatEtttc
13561 ccatEaatct aaattgcttt taaaagctca attgcaagat tagtatttaa atacattgag
    
```

アミノ酸配列: MGKIK 塩基配列: ATG GGA AAA