

H22 近畿大学・農学部・生命情報学

分子生物学の基礎と 配列データベース

2010年4月13日(火)

奈良先端大・情報・蛋白質機能予測学講座

川端 猛

takawaba@is.naist.jp

<http://isw3.naist.jp/IS/Kawabata-lab/home-ja.html>

平成22年度「生命情報学&生命情報学実習」講義日程

	講義	生命情報学	演習	生命情報学演習	2010.3.25
4/13	川端1	分子生物学の基礎と配列データベース			
4/20	川端2	ペアワイズアライメントと配列相同性検索	川端	主要WEBデータベースと配列相同性検索	
4/27	川端3	マルチプルアライメントとその応用			
5/11	川端4	分子系統学基礎	川端	マルチプルアライメントと系統樹作成演習	
5/18	川端5	蛋白質の物理化学的性質とアミノ酸配列解析			
5/25	川端6	蛋白質立体構造データの情報解析	川端	蛋白質立体構造データの可視化	
6/1	川端	>>試験(川端 担当分)<<			
6/8	中村1	化学構造データと計算化学基礎I			
6/15	中村2	化学構造データと計算化学基礎II	中村	ChemOfficeを用いた計算化学演習	
6/22	中村	>>試験(中村担当分)<<			
6/29	金谷1	トランスクリプトーム解析			
7/6	金谷2	インタラクローム解析	金谷	発現プロフィール解析演習	
7/13	金谷3	メタボローム解析	金谷	インタラクローム・代謝物解析演習	
7/20	金谷	>>試験(金谷担当分)<<			

講義の進め方について(川端)

- 毎回、出席を兼ねて小テストをします。
- 講義のスライドのPDFファイルは、講義 <http://isw3.naist.jp/IS/Kawabata-lab/lec-ja.html>のサイトに置きます。
- 川端分の試験は6/1(火)に行います。筆記用具以外持ち込み不可とします。
- 4/20(火),5/11(火),5/25(火)の午後には、講義と対応する内容の「生命情報学演習」を行います。

全ゲノム配列が決定された生物種

Apr 9, 2010

生物種		完了		ドラフト配列	進行中
原核生物	古細菌	66	メタン菌、超好熱菌、高度好塩菌など	7	26
	真正細菌	659	大腸菌、乳酸菌、コレラ菌、結核菌、シアノバクテリアなど	531	475
真核生物	動物	7	ヒト、チンパンジー、ブタ、カ、ショウジョウバエ2種、線虫	102	113
	植物	6	シロイヌナズナ、コメ、トウモロコシ、緑藻3種	18	69
	真菌	17	出芽酵母、分裂酵母、カンジダなど	78	34
	原生生物	9	マラリア原虫、赤痢アメーバなど	29	35
合計		763		767	756

<http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html>から転載

全ゲノムが解読された主な生物種(発表年代順)

発表年	生物種	ゲノムサイズ (M(10^6)塩基対)	遺伝子数
1995	マイコプラズマ菌 (<i>Mycoplasma genitalium</i>)	0.6	467
	インフルエンザ菌 (<i>Haemophilus influenzae</i>)	1.8	1717
1997	出芽酵母 (<i>Saccharomyces cerevisiae</i>)	12.1	6140
	大腸菌(<i>Escherichia coli</i>)	4.6	4289
1998	線虫(<i>Caenorhabditis elegans</i>)	97.0	19099
2002	マウス(<i>Mus musculus</i>)	2625.0	25865
2003	ヒト(<i>Homo sapiens</i>)	3068.0	26626

- ・一番小さいマイコプラズマでも $0.6 \times 10^6 = 60$ 万文字の{A,T,G,C}
- ・フロッピーディスク: 1.2M, CD-ROM: 600M, DVD: 4000Mなので、
バクテリアゲノムはフロッピー数枚、ヒトゲノムはDVDに収納可能

ゲノムデータを収納するには？



大腸菌のゲノム

4.6M塩基対

フロッピーディスク (1.2MB) 4枚分



ヒトのゲノム

3.0G塩基対

DVDディスク (4.7GB) 1枚分

今日の講義の内容

分子生物学で扱うデータ(DNA配列、アミノ酸配列)について

(1) そもそもDNAとは？ 蛋白質とは？

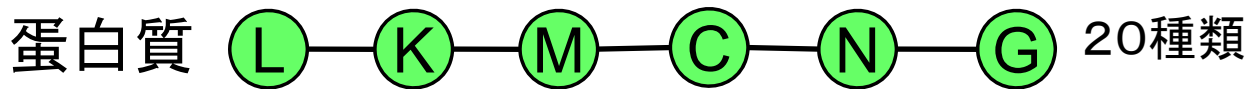
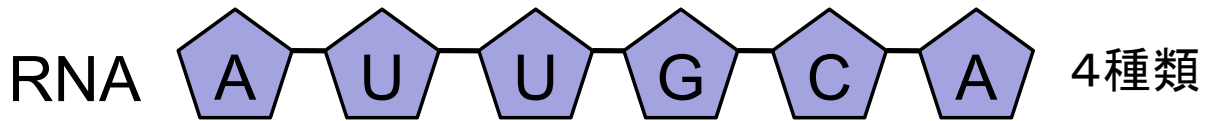
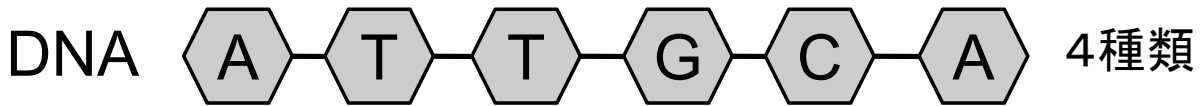
(2) どんなデータベースに、どのように収納されているか？

分子生物学の基礎

DNA→RNA→蛋白質の情報の流れ

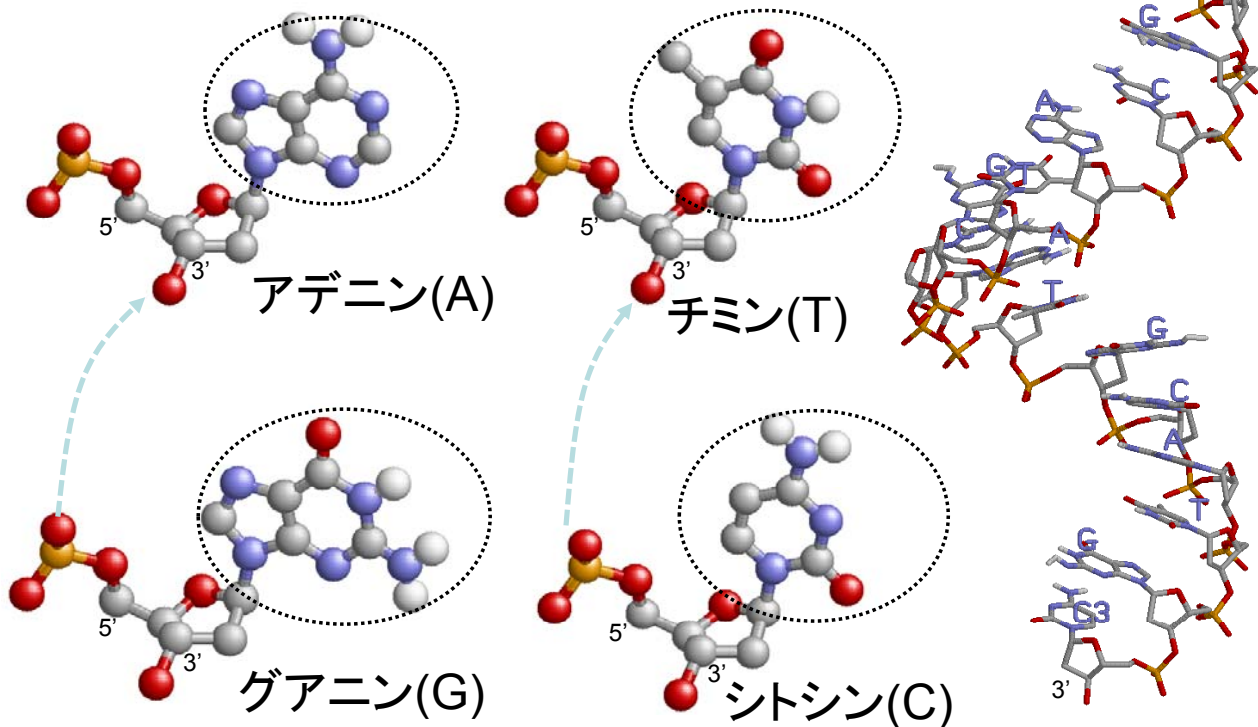
三つの重要な高分子 ～ DNA, RNA, 蛋白質 ～

これら三つはいずれも重合体(polymer)、つまりある単位となる分子(monomer)が一行に並んだ形



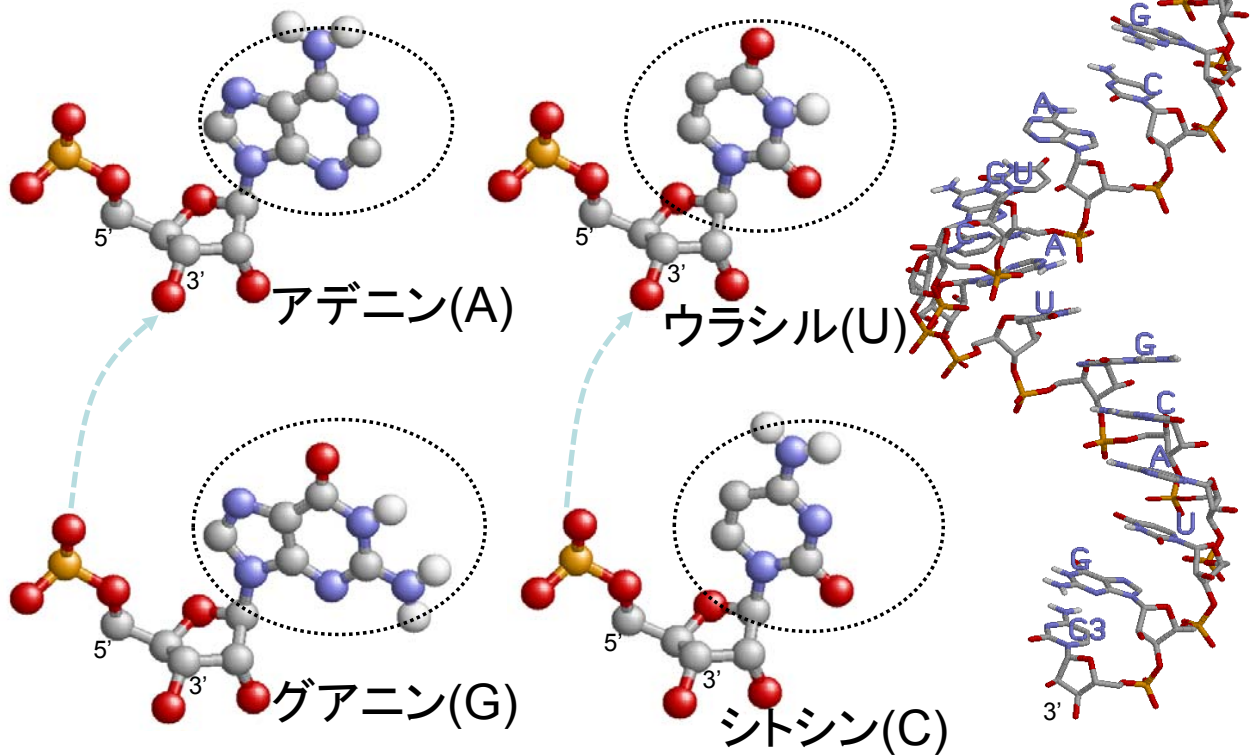
DNAの構成要素

4種のヌクレオチドでデオキシリボ核酸を構成

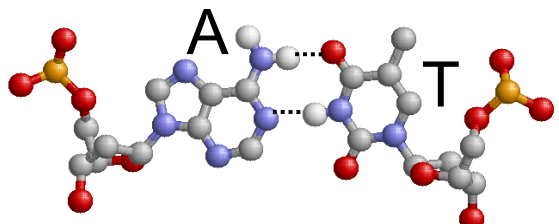


RNAの構成要素

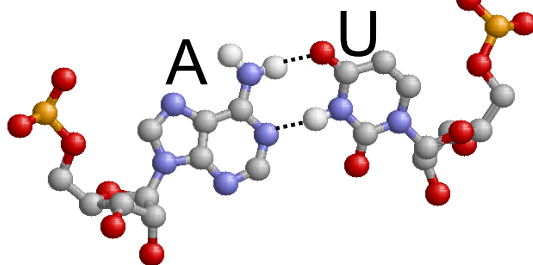
4種のヌクレオチドでリボ核酸を構成



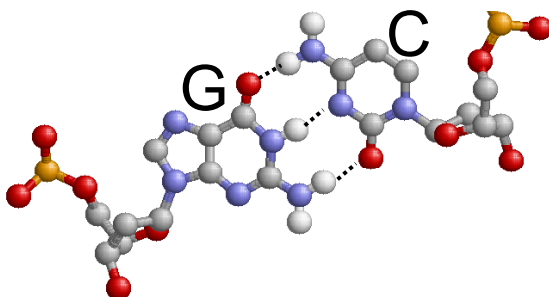
相補的な塩基対の構造



・向かいあう塩基どうしが水素結合を作る。
これを「相補的な塩基対」と呼ぶ。

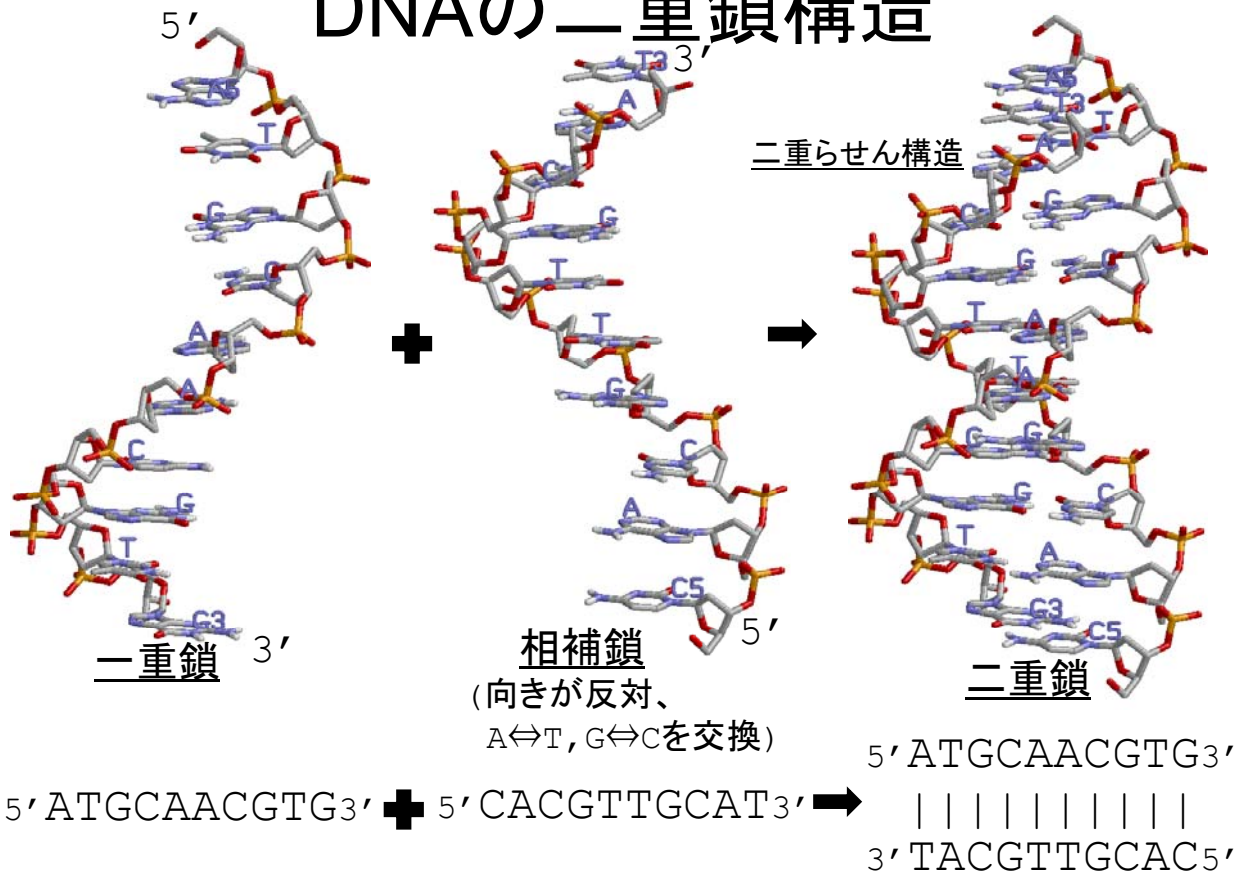


・A-T, A-U, G-Cの三種のペアが可能

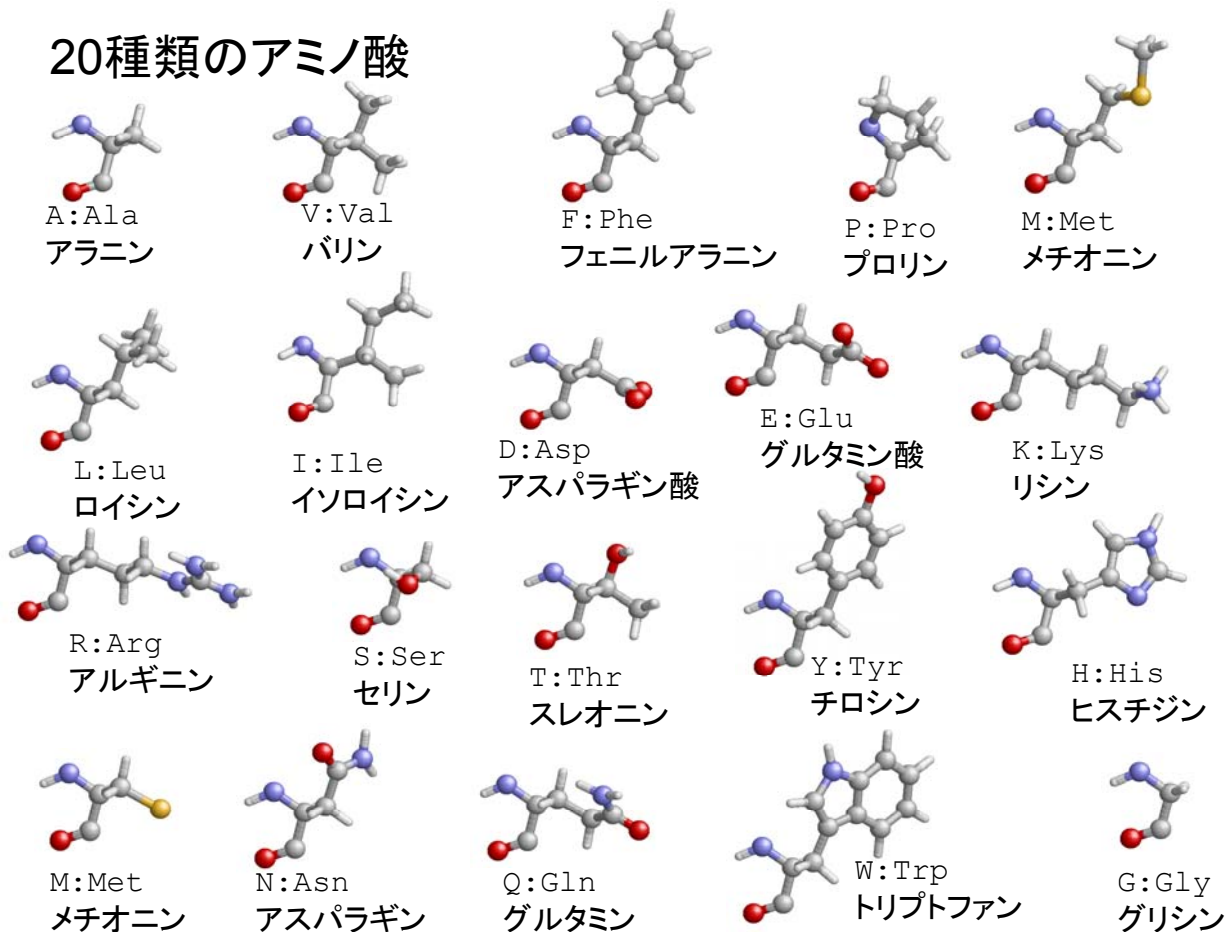


・DNAどうし、RNAどうしだけでなく、DNAとRNAのペアも可能

DNAの二重鎖構造

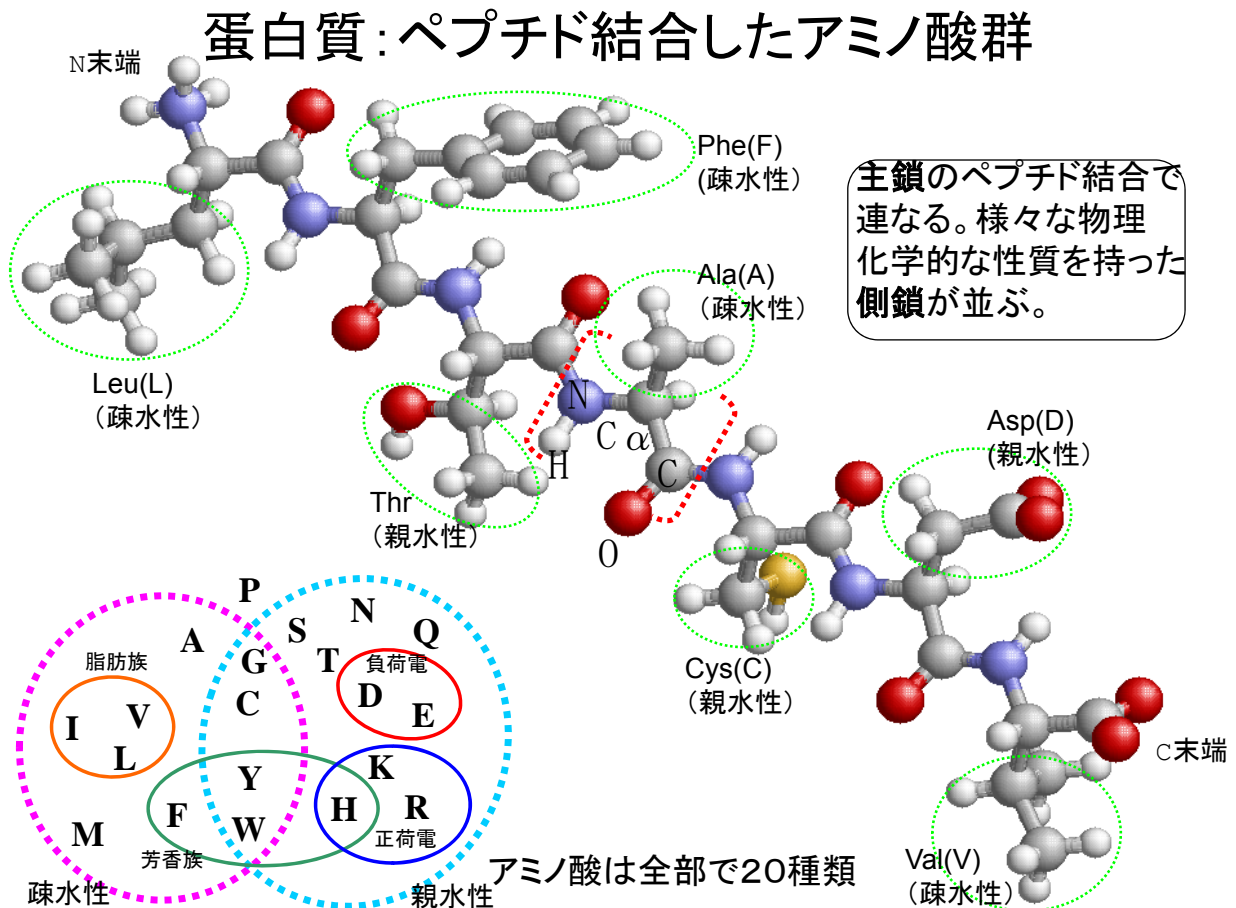


20種類のアミノ酸

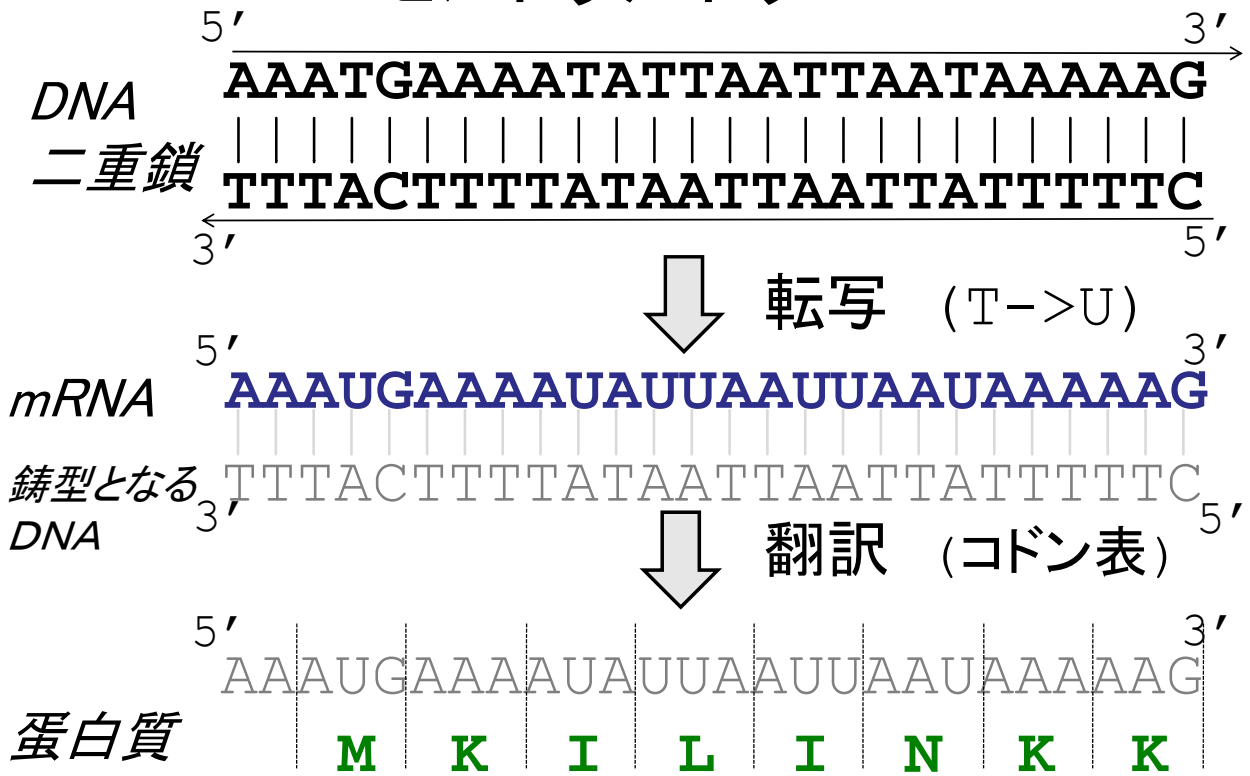


アミノ酸の一文字表記を覚えましょう

- アラニン(A)
- ロイシン(L)
- フェニルアラニン(F)
- トリプトファン(W)
- リジン(K)
- グルタミン(Q)
- グルタミン酸(E)
- アスパラギン(N)
- アスパラギン酸(D)



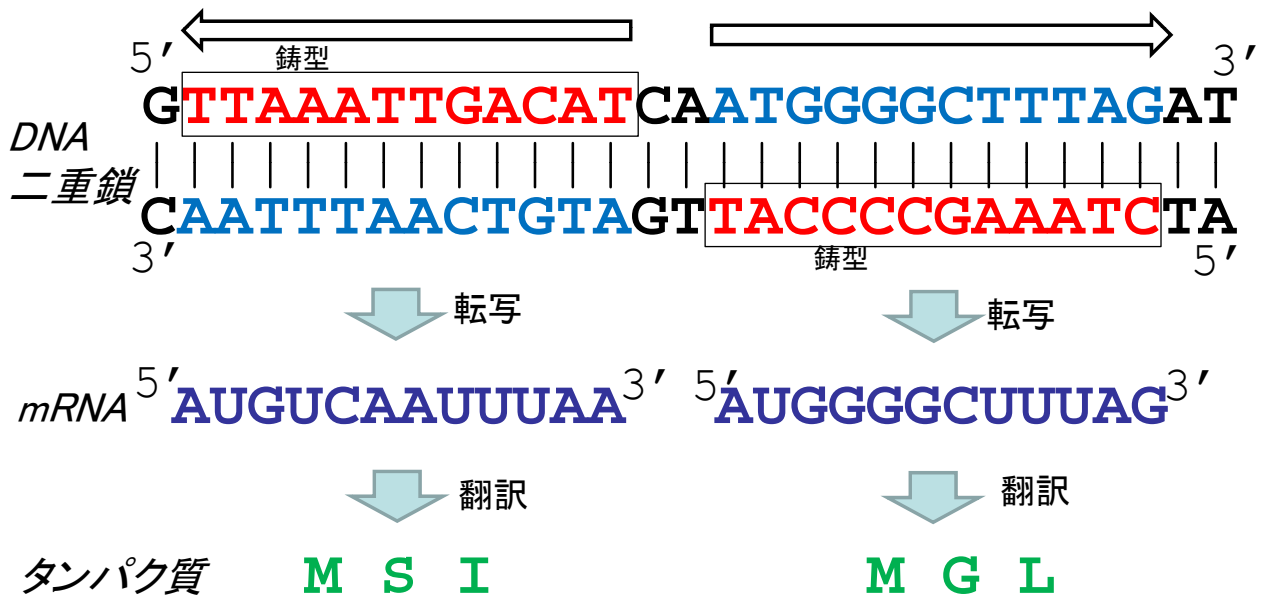
セントラルドグマ



遺伝暗号(コドン表):RNA

UUU	F:Phe	UCU	S:Ser	UAU	Y:Tyr	UGU	C:Cys
UUC		UCC		UAC		UGC	
UUA	L:Leu	UCA		UAA	終止	UGA	終止
UUG		UCG		UAG		UGG	W:Trp
CUU	L:Leu	CCU	P:Pro	CAU	H:His	CGU	R:Arg
CUC		CCC		CAC		CGC	
CUA		CCA		CAA	Q:Gln	CGA	
CUG		CCG		CAG		CGG	
AUU	I:Ile	ACU	T:Thr	AAU	N:Asn	AGU	S:Ser
AUC		ACC		AAC		AGC	
AUA		ACA		AAA	K:Lys	AGA	R:Arg
AUG	M:Met(開始)	ACG		AAG		AGG	
GUU	V:Val	GCU	A:Ala	GAU	D:Asp	GGU	G:Gly
GUC		GCC		GAC		GGC	
GUA		GCA		GAA	E:Glu	GGA	
GUG		GCG		GAG		GGG	

DNAの二重鎖のどちらを鋳型にしてもよい

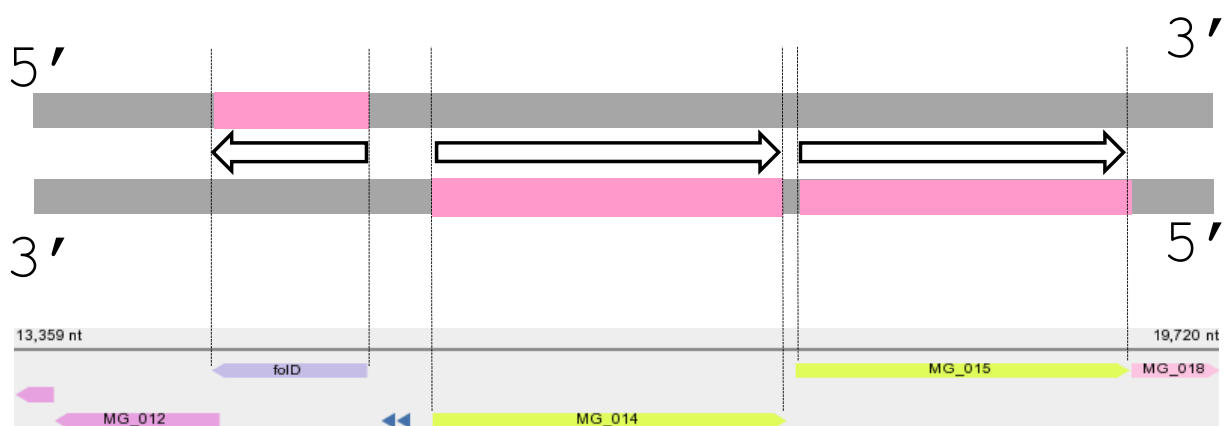


※DNA配列のデータベースでは5' ⇒ 3' の方向に塩基配列が記載。

その配列を 鋳型として転写された遺伝子 : 右から左 ←

反対側の配列を鋳型として転写された遺伝子 : 左から右 →

どちらを鋳型にしたかを矢印で示す



*Mycoplasma genitalium*のゲノムの一部

※DNA配列のデータベースでは5' ⇒ 3' の方向に塩基配列が記載。

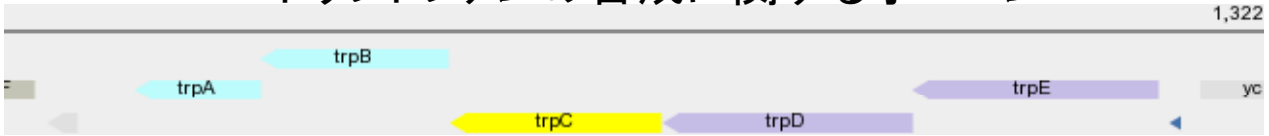
その配列を 鋳型として転写された遺伝子 : 右から左 ←

反対側の配列を鋳型として転写された遺伝子 : 左から右 →

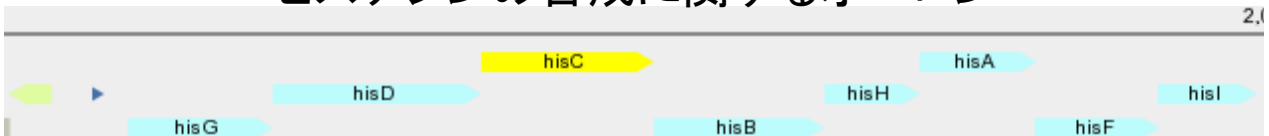
バクテリアのオペロン構造

一つの機能を担う遺伝子群が隣接して同じ方向にコードされた領域

トリプトファンの合成に関するオペロン



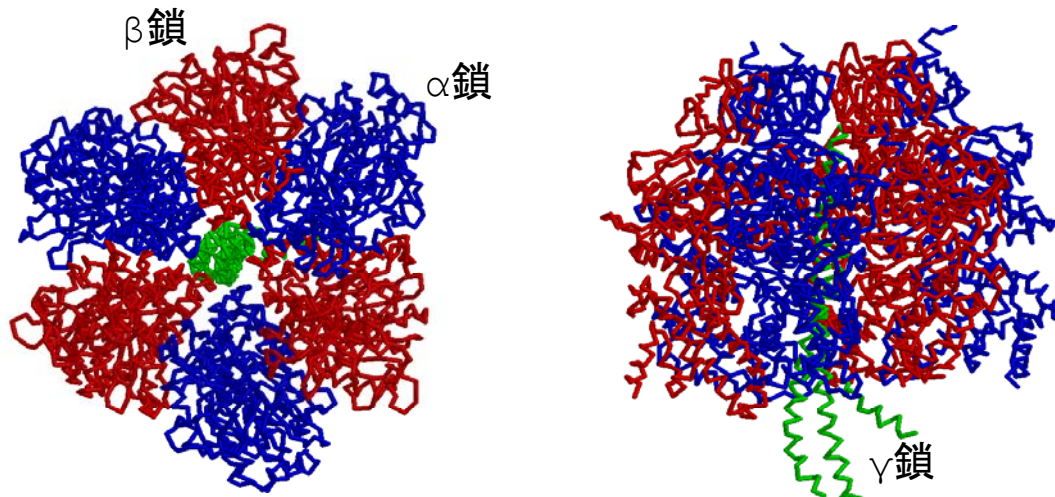
ヒスチジンの合成に関するオペロン



いくつかの遺伝子をコードする領域がまとめてmRNAに転写され、まとめて、タンパク質に翻訳される。

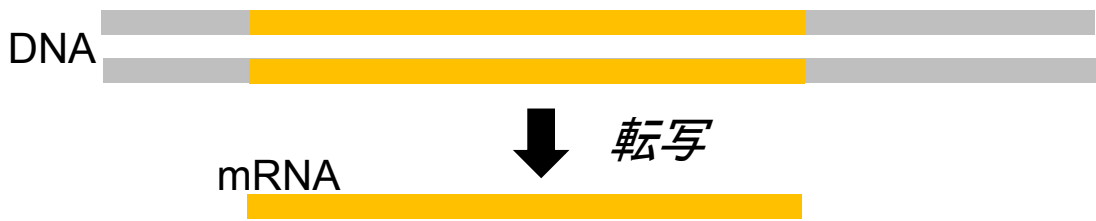
オペロンにコードされた遺伝子の 蛋白質が複合体を構成する例

F₁F₀ATP合成酵素のオペロン

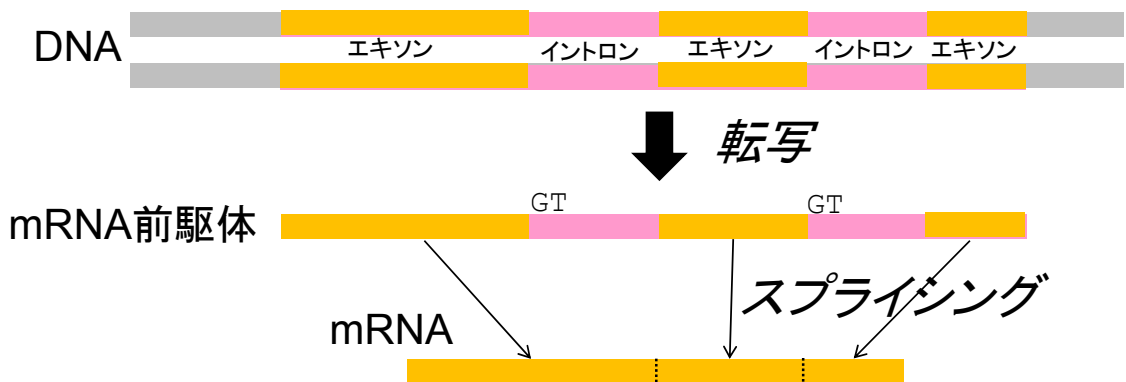


真核生物はエクソン・イントロン構造を持つ

・原核生物(prokaryote)の場合



・真核生物(eukaryote)の場合



mRNAの翻訳の例

mRNA

AGCAAUGAAAUAUUAAUUAAUAAAUAACGA

(1) まず開始コドンのAUGを探す

AGCAAUGAAAUAUUAAUUAAUAAAUAACGA
→

(2) そのまま3文字ずつスライドしてコドン表に従って翻訳

AGCAAUGAAAAUAUUAAUUAAUAAAUAACGA
M K I
→

(3) 終始コドン (UAA, UAG, UGA) が現れたら終了

AGCAAUGAAAAUAUUAAUUAAUAAAUAACGA
M K I L I N K 終止
→

翻訳に関係するいくつかの用語

gene (遺伝子) : 生物学的情報を含んでいるDNAの部分領域であり、RNAあるいは蛋白質をコードする部分。

CDS : CoDing Sequenceの略。蛋白質をコードしている核酸配列の領域。

ORF : Open Reading Frameの略。開始コドンから始まり、終止コドンで終わる核酸配列の領域

DNA配列からアミノ酸配列を予測できるか？

イントロンのない原核生物の場合

- ・6通りの読み枠(reading frame)を全て試し、
- ・開始コドンで始まり終止コドンで終わる領域(open reading frame)を抽出
- ・十分長い領域を翻訳されるアミノ酸配列として予測

AGCAAUGAAAUAUUAUUAAUAAAUAAC

S N E N I N x x I

A M K I L I N K x

Q x K Y x L I N N

※一つの方向あたり三つの読み枠がある。
相補鎖にも三つあるので、全部で6つの読み枠。

遺伝暗号(コドン表):DNA

TTT	F:Phe	TCT	S:Ser	TAT	Y:Tyr	TGT	C:Cys
TTC		TCC		TAC		TGC	
TTA	L:Leu	TCA		TAA	終止	TGA	終止
TTG		TCG		TAG		TGG	W:Trp
CTT	L:Leu	CCT	P:Pro	CAT	H:His	CGT	R:Arg
CTC		CCC		CAC		CGC	
CTA		CCA		CAA	Q:Gln	CGA	
CTG		CCG		CAG	CGG		
ATT	I:Ile	ACT	T:Thr	AAT	N:Asn	AGT	S:Ser
ATC		ACC		AAC		AGC	
ATA		ACA		AAA	K:Lys	AGA	R:Arg
ATG	M:Met(開始)	ACG		AAG		AGG	
GTT	V:Val	GCT	A:Ala	GAT	D:Asp	GGT	G:Gly
GTC		GCC		GAC		GGC	
GTA		GCA		GAA	E:Glu	GGA	
GTG		GCG		GAG		GGG	

配列決定とバイオインフォマティクス: 学籍番号: _____ 氏名: _____

問1. 以下のDNA配列の3つの読み枠について、それぞれ対応するアミノ酸を1文字表記で記せ。終止コドンは'x'と書け。

GATGAATGTATTTGCCTGAGTCCTTTCTGAAA

GATGAATGTATTTGCCTGAGTCCTTTCTGAAA

GATGAATGTATTTGCCTGAGTCCTTTCTGAAA

問2. 最も長いORFに対応するアミノ酸配列は何か。以下に記せ。

アミノ酸配列: _____

問1. 以下のDNA配列の3つの読み枠について、それぞれ対応するアミノ酸を1文字表記で記せ。終止コドンは'x'と書け。

GATGAATGTATTTGCCTGAGTCTTTCTGAAA
D E C I C L S L S E

GATGAATGTATTTGCCTGAGTCTTTCTGAAA
M N V F A x V F L K

GATGAATGTATTTGCCTGAGTCTTTCTGAAA
x M Y L P E S F x

問2. 最も長いORFに対応するアミノ酸配列は何か。以下に記せ。

アミノ酸配列: **MYLPESF**

H22 情報生命学:2010. 4. 13

より正確に遺伝子を予測するには？

専用の遺伝子予測プログラムの使用が推奨
(GeneHacker, GeneMark, Glimmer)

- 開始コドンの前の配列の特徴
- 遺伝子領域の塩基配列の規則性

GeneHacker
- A System for Gene Structure Prediction in Microbial Genomes -

解析条件

解析対象配列
配列データ長は50000までです。
 配列データファイルのアップロード (認識可能なファイル形式: *.txt/*gz/*Z/*lzh/*zip) [参照]
 直接入力

パラメーター

Species: Bacillus subtilis (selections)
Strand: Direct (selections)
Overlapping genes: Yes (selections)
Frame-shift error rate: 0.0 (0.0 to 100.0)

Eメールアドレス: _____ (ジョブ終了の通知を受け取りたい方は入力してください)

真核生物の遺伝子の予測



エキソン部分だけうまく抽出する必要があり、大変難しい

- ・真核生物用の遺伝子発見プログラムも開発されている (Genscan, HMMgene, Grail II, GeneParser)
- ・mRNAのデータ (cDNAやEST) の利用が手堅い
- ・既知の遺伝子との類似領域の比較 (blastxなど) も有効

配列データベースの成り立ち

DNA, RNA, 蛋白質の配列を決める実験法

⇒直接配列を計測できるのはDNAだけ

- DNA配列はPCR技術を用いて、注目領域を増幅し、ジデオキシ法を用いたDNAシーケンサを使って決定することができる。
- RNA配列は、RNAをDNAに逆転写し、そのDNA配列を決定することで、得ることができる。
- 蛋白質からそのアミノ酸配列を得るのは難しく、特に全長を決めるのは困難。DNAかRNAの配列を解釈することでアミノ酸配列を得ることが一般的。

国際塩基配列データベース



日本：DDBJ（遺伝研）
米国：GenBank（NCBI）
欧州：EMBL-Bank（EBI）

研究者は決定したDNA配列を三つのデータベースのどれかに登録

どのデータベースに登録しても、データは共有される

GenBankとDDBJはgenbank形式と呼ばれるファイルフォーマットを採用

NCBIのWEBページの例

マイコプラズマ菌の全ゲノム配列のトップページ

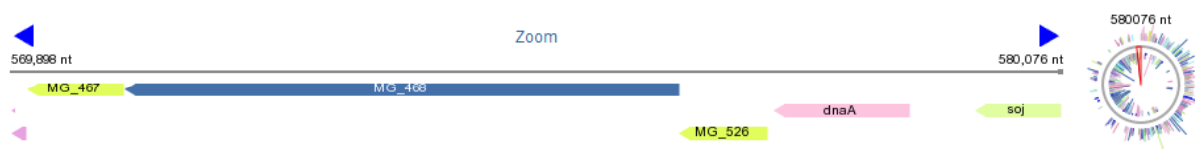
[Genome](#) > [Bacteria](#) > [Mycoplasma genitalium G37, complete genome](#)

Lineage: [Bacteria](#); [Tenericutes](#); [Mollicutes](#); [Mycoplasmatales](#); [Mycoplasmataceae](#); [Mycoplasma](#); [Mycoplasma genitalium](#); [Mycoplasma genitalium G37](#)

Genome Info:	Features:	BLAST homologs:	Links:	Review Info:
Refseq: NC_000908	Genes: 524	COG	Genome Project	Publications: [4]
GenBank: L43967	Protein coding: 475	TaxMap	Refseq FTP	Refseq Status: PROVISIONAL
Length: 580,076 nt	Structural RNAs: 43	TaxPlot	GenBank FTP	Seq. Status: Completed
GC Content: 31%	Pseudo genes: 6	GenePlot	BLAST	Sequencing center: TIGR
% Coding: 90%	Others: 6	gMap	TraceAssembly	Completed: 2001/01/08
Topology: circular	Contigs: None		CDD	Organism Group
Molecule: dsDNA			Other genomes for species:	

Gene Classification based on [COG functional categories](#)

Search gene, GeneID or locus_tag:



マイコプラズマ菌の全ゲノムファイル:ヘッダー部

```

LOCUS       NC_000908                580076 bp    DNA    circular BCT 02-FEB-2009
DEFINITION  Mycoplasma genitalium G37, complete genome.
ACCESSION   NC_000908
VERSION     NC_000908.2   GI:108885074
KEYWORDS    .
SOURCE      Mycoplasma genitalium G37
  ORGANISM  Mycoplasma genitalium G37
             Bacteria; Tenericutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
REFERENCE   1  (bases 1 to 580076)
AUTHORS     Glass,J.I., Assad-Garcia,N., Alperovich,N., Yooseph,S., Lewis,M.R.,
             Maruf,M., Hutchison,C.A., Smith,H.O. and Venter,J.C.
TITLE       Essential genes of a minimal bacterium
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 103 (2), 425-430 (2006)
PUBMED     16407165
REFERENCE   2  (bases 1 to 580076)
AUTHORS     Peterson,S.N., Bailey,C.C., Jensen,J.S., Borre,M.B., King,E.S.,
             Bott,K.F. and Hutchison,C.A.III.
TITLE       Characterization of repetitive DNA in the Mycoplasma genitalium
             genome: possible role in the generation of antigenic variation
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 92 (25), 11829-11833 (1995)
PUBMED     8524858
REFERENCE   3  (bases 1 to 580076)
AUTHORS     Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A.,
             Fleischmann,R.D., Bult,C.J., Kerlavage,A.R., Sutton,G.G.,
             Kelley,J.M., Fritchman,J.L., Weidman,J.F., Small,K.V., Sandusky,M.,
             Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Saudek,D.M.,
             Phillips,C.A., Merrick,J.M., Tomb,J., Dougherty,B.A., Bott,K.F.,

```

マイコプラズマ菌の全ゲノムファイル:FEATURES(1)

FEATURES	Location/Qualifiers
gene	686..1828 /gene="dnaN" /locus_tag="MG_001" /db_xref="GeneID:875454"
CDS	686..1828 /gene="dnaN" /locus_tag="MG_001" /EC_number="2.7.7.7" /note="identified by sequence similarity; putative" /codon_start=1 /transl_table=4 /product="DNA polymerase III, beta subunit" /protein_id="NP_072661.2" /db_xref="GI:108885075" /db_xref="GeneID:875454" /translation="MKILINKSELNKKILKMNVIISNNKIKPHHSYFLIEAKEKEIN FYANNEYFSVKCNLKNIDILEQGSLLIVKGIKIFNDLINGIKEEIIITIQEKDQTLVKT KKTSLNLTINVNEFPRIREFNEKNDLSEFNQFKINYSLLVKGKIKIFHVSNNREISS KFNGVNFNGSNGKEIFLEASDTYKLSVFEIKQETEPDFILESNNLSFINSFNPEEDK SIVFYRDKDNKDSFSTEMLISMDNFMISYTSVNEKFPVNYFFEFEPETKIVVQKNEL KDALQRIQTLAQNERTFLCDMQINSSELKIRAIVNNIGNSLEEISCLKFEGYKLNISF NPSSLLDHIESFESNEINFDFQGNISKYFLITSKSEPELQKQILVPSR"
gene	1828..2760 /locus_tag="MG_002" /db_xref="GeneID:875221"
CDS	1828..2760

マイコプラズマ菌の全ゲノムファイル:FEATURES(2)

gene	complement(12701..13564) /locus_tag="MG_011"
CDS	complement(12701..13564) /locus_tag="MG_011" /note="identified by sequence similarity; putative" /codon_start=1 /product="hypothetical protein" /protein_id="NP_072671.1" /translation="MGKIKLKNRKALVVYDNKDDFEKNQTFALSLLIKELQKKKLNAEV LLENKDINFREAKINEAELILNRSRKVDFLKTNNQINTFLVNPFPVNVFIANDKYETYK WLKQNRFLTIVNSLLSKETIKSFPVIVKRNSHGGKDVHLVNSADEIKHLNIENATEW IVQPFLSIGTVEYRAYILFGKIIKVIKKISNANQFKANFSQGAEVSLFKLKWFTKRKI KKIAKRLREGYYAIDFFLNRYNRVIVNEIEDAAGARALVQLCPDLNITKIIIRTIISK FKKFLKKKLIS"
gene	complement(15294..15369) /locus_tag="MG_471" /note="MG_t01" /db_xref="GeneID:875702"
tRNA	complement(15294..15369) /locus_tag="MG_471" /product="tRNA-Ala"
gene	complement(15375..15451) /locus_tag="MG_472" /old_locus_tag="MGt02" /db_xref="GeneID:875218"
tRNA	complement(15375..15451) /locus_tag="MG_472"

マイコプラズマ菌の全ゲノムファイル:塩基配列(1)

ORIGIN

```
1 taagttatta tttagttaat acttttaaca atattattaa ggtatttaaa aaatactatt
61 atagtatatta acatagttaa ataccttcct taatactggt aaattatatt caatcaatac
121 atatataata ttattaaaat acttgataag tattatttag atattagaca aatactaatt
181 ttatattgct ttaataactta ataaatacta cttatgtatt aagtaaatat tactgtaata
241 ctaataacaa tattattaca atatgctaga ataattatgc tagtatcaat aattactaat
301 atagtattag gaaaatacca taataatatt tctacataat actaagttaa tactatgtgt
361 agaataataa ataatcagat taaaaaaatt ttatttatct gaaacatatt taatcaattg
421 aactgattat tttcagcagt aataattaca tatgtacata gtacatatgt aaaatatcat
481 taatttctgt tatatataat agtatctatt ttagagagta ttaattatta ctataattaa
541 gcatttatgc ttaattataa gctttttatg aacaaaatta tagacatttt agttcttata
601 ataaataata gatattaaag aaaataaaaa aatagaaaata aatatcataa cccttgataa
661 cccagaaatt aatacttaat caaaaatgaa aatattaatt aataaaaagt aattgaataa
721 aattttgaaa aaaatgaata acgtttattat ttccaataac aaaataaaac cacatcattc
781 atatttttta atagaggcaa aagaaaaaga aataaaacttt tatgctaaca atgaatactt
841 ttctgtcaaa tgtaatttaa ataaaaatat tgatattcct gaacaaggct ccttaattgt
901 taaaggaaaa atttttaacg atcttattaa tggcataaaa gaagagatta ttactattca
961 agaaaaagat caaacacttt tggttaaaac aaaaaaaca agtattaatt taacacaat
1021 taatgtgaat gaatttccaa gaataagggt taatgaaaa aacgatttaa gtgaatttaa
1081 tcaattcaaa ataaattatt cacttttagt aaaaggcatt aaaaaaattt ttcactcagt
1141 ttcaaaaat cgtgaaatat cttctaaatt taatggagta aatttcaatg gatccaatgg
1201 aaaagaataa tttttagaag cttctgcac ttataaacta tctgtttttg agataaaagca
1261 agaaacagaa ccatttgatt ccttttgga gtagaattta cttagtttca ttaattcttt
1321 taatcctgaa gaagataaat ctattgtttt ttattacaga aaagataata aagatagctt
1381 tagtacagaa atggtgattt caatggataa ctttatgatt agttacacat cgttaatga
1441 aaaatttcca gaggtaaact acttttttga atttgaacct gaaactaaaa tagttgttca
```

:

マイコプラズマ菌の全ゲノムファイル塩基配列(2)

```
:
12601 aactaagcaa ggatttataa caaaagttat agaaattaa gctgccgcaa aagactgaaa
12661 tgatttgttt ttattaaaca actcaaattg atcagcgggt ttaactaatc aacttctttt
12721 ttaagaattt tttaaattta ctaataattg ttctgataat tatttttagt atatttaa
12781 ctggacaaaag ctgaactaaa gctctcgcac cagcagcatc ttcaatttca ttaacaataa
12841 ccctattata tctattttaa aagaagtcaa tagcataata acctccctt aggcgttag
12901 ctattttctt tatttttctt ttagtaaadc actttaattt aaacaaggaa acttcagcac
12961 cttgtgaaaa gttagcttta aattgattag cattagaat ttttttaata actttaatta
13021 tttttccaaa caaaatataa gcacgatatt caactgtgcc aattgataaa aaaggttgaa
13081 caattcattc tgttgcattt tcaatgttta aatgtttgat ctcgtcagca ctattaacta
13141 aatgtacatc ttttccaccg tgtgaattac gtttcttaac gatgacagga aatgatttga
13201 ttgtttcttt actaagaaga gaagaattga cagttagaaa tctattttgt tttaatcatt
13261 tatatgtttc gtatttatcg tttgctataa aaacaacatt aaaaggatta actaaaaaag
13321 tattttattg attattgggt tttaaaaaat ctacttttct tgaacgattt aaaatcaatt
13381 cagcttcatt aatttttagct tcgaaattaa tgcctttatt ttcaagtaat aagacttcag
13441 catttagttt tttcttttgt aattccttga ttagacttaa agcaaagtgt tgattttttt
13501 caaaatcatc cttgttgtca taacaacta atgcttttct gtttttaata ttaatttttc
13561 ccattaatct aaattgcttt taaaagctca attgcaagat tagtatttaa atacattgag
13621 cttcttggtt attgcacatt aggatttact tcacaaaaga tcaatgatct gtcttgatca
13681 aacaaaaaat caataccgca ataaaaaagt tgcattactt tactaatttt aactgctaaa
13741 ttttcttggt ctttattcaa aaaaaagcgt tctgccttgg cccctttatt gagattagaa
13801 cgaaaatcac tattattagt tgtatgtaa gcacctataa ctttattggt cacaacaata
:
579961 atgatcctgc aacattagtt gccattgtag tttttaatac gccgccttta ttatttacia
580021 aagaaatgat catatattta aatgattata atatttcttt aactactaaa aaatac
```

//

核酸配列に付加される主なFEATURE

gene : 遺伝子

CDS : Coding Sequence

tRNA : transfer RNA(運搬RNA)

※配列情報以外に付加される情報のことをアノテーション(annotation)と呼ぶ。

FEATUREの領域の書き方

```
CDS 1828..2760
```

1828～2760番目の塩基配列 (その相補鎖を鋳型として転写されたRNA)

```
CDS complement(1807..2169)
```

1807～2169番目の相補鎖の塩基配列 (その配列を鋳型として転写されたRNA)

```
CDS join(7287..7388,7502..7753)
```

7287～7388番目と7502～7733番目の配列を加えた配列
(複数のエキソンからなる遺伝子の記述に用いる)

```
CDS complement(join(7287..7388,7502..7753))
```

7287～7388番目と7502～7733番目の配列を加えた配列
の相補鎖の配列。
(複数のエキソンからなる遺伝子が相補鎖にある場合)

問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

- (1) 遺伝子 dnaN の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : _____
アミノ酸 : _____

- (2) 遺伝子 MG_011 の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : _____
アミノ酸 : _____

問4. ヒトの DNA 配列のエントリ (NC_000012) のデータを見て以下の間に答えよ。

- (1) 遺伝子 TPI1 が mRNA に転写される領域の最初と最後の DNA 4 文字を書け

最初: _____ 最後: _____

- (2) 遺伝子 TPI1 を構成するエキソンの数を記せ。 _____ 個

- (3) 遺伝子 TPI1 の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の 3 塩基を書け。

第1領域の開始 | _____ 終結 | _____
_____ | _____ _____ | _____
第2領域の開始 | _____ 終結 | _____
_____ | _____ _____ | _____

マイコプラズマ菌の全ゲノムファイル: FEATURES(1)

```
FEATURES             Location/Qualifiers
     gene             686..1828
                     /gene="dnaN"
                     /locus_tag="MG_001"
                     /db_xref="GeneID:875454"
     CDS              686..1828
                     /gene="dnaN"
                     /locus_tag="MG_001"
                     /EC_number="2.7.7.7"
                     /note="identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=4
                     /product="DNA polymerase III, beta subunit"
                     /protein_id="NP_072661.2"
                     /db_xref="GI:108885075"
                     /db_xref="GeneID:875454"
                     /translation="MKILINKSELNKKILKKMNNVLIISNNKIKPHHSYFLIEAKEKEIN
FYANNEYFSVKCNLKNIDILEQGS LIVKGKIFNDLINGIKEEIIITIQEKDQTL LVKT
KKTSINLNTINVNEFPRIRFNEKNDLSEFNQFKINYSLLVKGKIKKIFH SVSNNREISS
KFNGVNFNGSNGKEIFLEASD TYKLSVFEIKQETEPDFDFILES NLLSFINSFNPEEDK
SIVFYRKNKDSFSTEMLISMDNFMISYTSVNEKFP EVNYFFEFEPETKIVVQKNEL
KDALQRIQTLAQNERTFLCDMQINSSELKIRAI VNNIGNSLEEISCLKFEGYKLNISF
NPSSLLDHIESFESNEINFDFQGN SKYFLITSKSEPELQILVPSR"
     gene             1828..2760
                     /locus_tag="MG_002"
                     /db_xref="GeneID:875221"
     CDS              1828..2760
```

マイコプラズマ菌の全ゲノムファイル:塩基配列(1)

ORIGIN

```
1 taagttatta tttagttaat acttttaaca atattattaa ggtatttaaa aaatactatt
61 atagtatatta acatagttaa ataccttcct taatactggt aaattatatt caatcaatac
121 atatataata ttattaaaat acttgataag tattatttag atattagaca aatactaatt
181 ttatattgct ttaataactta ataaatacta cttatgtatt aagtaaatat tactgtaata
241 ctaataacaa tattattaca atatgctaga ataattatgc tagtatcaat aattactaat
301 atagtattag gaaaatacca taataatatt tctacataat actaagttaa tactatgtgt
361 agaataataa ataatcagat taaaaaaatt ttatttatct gaaacatatt taatcaattg
421 aactgattat tttcagcagt aataattaca tatgtacata gtacatatgt aaaatatcat
481 taatttctgt tatatataat agtatctatt ttagagagta ttaattatta ctataattaa
541 gcatttatgc ttaattataa gctttttatg aacaaaatta tagacatttt agttcttata
601 ataaataata gatattaaag aaaataaaaa aatagaaaata aatatcataa cccttgataa
661 cccagaaatt aatacttaat caaaaatgaa aatattaatt aataaaaagt aattgaataa
721 aattttgaaa aaaatgaata acgttattat ttccaataac aaaataaaac cacatcattc
781 atatttttta atagaggcaa aagaaaaaga aataaaacttt tatgctaaca atgaatactt
841 ttctgtcaaa tgtaatttaa ataaaaatat tgatattcct gaacaaggct ccttaattgt
901 taaaggaaaa atttttaacg atcttattaa tggcataaaa gaagagatta ttactattca
961 agaaaaagat caaacacttt tggttaaaac aaaaaaaca agtattaatt taacacaat
1021 taatgtgaat gaatttcaa gaataagggt taatgaaaa aacgatttaa gtgaatttaa
1081 tcaattcaa ataaattatt cacttttagt aaaaggcatt aaaaaaattt ttcactcagt
1141 ttcaaataat cgtgaaatat cttctaaatt taatggagta aatttcaatg gatccaatgg
1201 aaaaagaata tttttagaag cttctgacac ttataaacta tctgttttg agataaagca
1261 agaaacagaa ccatttgatt tcattttgga gagtaattta cttagtttca ttaattcttt
1321 taatcctgaa gaagataaat ctattgtttt ttattacaga aaagataata aagatagctt
1381 tagtacagaa atggtgattt caatggataa ctttatgatt agttacacat cgttaatga
1441 aaaatttcca gaggtaaact acttttttga atttgaacct gaaactaaa tagttgttca
:
```

問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

- (1) 遺伝子 dnaN の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : **a t g a a a a t a**
アミノ酸 : _____

- (2) 遺伝子 MG_011 の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : _____
アミノ酸 : _____

問4. ヒトの DNA 配列のエントリ (NC_000012) のデータを見て以下の間に答えよ。

- (1) 遺伝子 TPI1 が mRNA に転写される領域の最初と最後の DNA 4 文字を書け

最初: _____ 最後: _____

- (2) 遺伝子 TPI1 を構成するエキソンの数を記せ。 _____ 個

- (3) 遺伝子 TPI1 の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の 3 塩基を書け。

第1領域の開始 | _____ 終結 | _____
_____ | _____ _____ | _____
第2領域の開始 | _____ 終結 | _____
_____ | _____ _____ | _____

問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

- (1) 遺伝子 dnaN の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : a t g a a a a t a
 アミノ酸 : M K I

- (2) 遺伝子 MG_011 の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : _____
 アミノ酸 : _____

問4. ヒトの DNA 配列のエントリ (NC_000012) のデータを見て以下の間に答えよ。

- (1) 遺伝子 TPI1 が mRNA に転写される領域の最初と最後の DNA 4 文字を書け

最初: _____ 最後: _____

- (2) 遺伝子 TPI1 を構成するエキソンの数を記せ。 _____ 個

- (3) 遺伝子 TPI1 の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の 3 塩基を書け。

第1領域の開始 | _____ 終結 | _____
 _____ | _____ _____ | _____
 第2領域の開始 | _____ 終結 | _____
 _____ | _____ _____ | _____

NCBI Sequence Viewerでの表示(dnaN)

The image shows the NCBI Sequence Viewer interface. The top panel displays the complete genome of *Mycoplasma genitalium* G37 (NC_000908) with a scale from 0 to 4 K bases. Gene models for dnaN (NP_072661.2), gyrB (NP_072663.1), and MG_002 (NP_072662.2) are shown. A blue dashed circle highlights the dnaN gene. The bottom panel provides a detailed view of the dnaN gene sequence from position 673 to 773. It shows the DNA sequence (ATCAAAAATCAAAAATATTAAATTAATAAACTCAATTG) and the amino acid sequence (M K I L I N K S E L) with their respective codons (ATG, AAA, ATA, TTA, ATT, AAT, AAA, AGT, GAA, TTG) listed below.

マイコプラズマ菌の全ゲノムファイル:FEATURES(2)

```
gene      complement (12701..13564)
          /locus_tag="MG_011"
CDS       complement (12701..13564)
          /locus_tag="MG_011"
          /note="identified by sequence similarity; putative"
          /codon_start=1
          /product="hypothetical protein"
          /protein_id="NP_072671.1"
          /translation="MGKIKLKNRKALVVYDNKDDFEKNQTFALSLIKELQKKKLNAEV
LLENKDINFREAKINEAELILNRSRKVDFLKTNNQINTFLVNPFNVVFIANDKYETYK
WLKQNRFLTVNSSLKETIKSFPVIVKKRNHGGKDVHLVNSADEIKHLNIENATEW
IVQPFLSIGTVEYRAYILFGKIIKVIKKISNANQFKANFSQGAEVSLFKLKWFTKRKI
KKIAKRLREGYYAIDFFLNRYNRVIVNEIEDAAGARALVQLCPDLNITKIIIRTIISK
FKKFLKKKLIS"
gene      complement (15294..15369)
          /locus_tag="MG_471"
          /note="MG_t01"
          /db_xref="GeneID:875702"
tRNA      complement (15294..15369)
          /locus_tag="MG_471"
          /product="tRNA-Ala"
gene      complement (15375..15451)
          /locus_tag="MG_472"
          /old_locus_tag="MGt02"
          /db_xref="GeneID:875218"
tRNA      complement (15375..15451)
          /locus_tag="MG_472"
```

マイコプラズマ菌の全ゲノムファイル塩基配列(2)

```
:
12601 aactaagcaa ggatttataa caaaagttat agaaattaa gctgccgcaa aagactgaaa
12661 tgatttgttt ttattaaaca actcaaattg atcagcggtt ttaactaatc aacttctttt
12721 ttaagaattt tttaaattta ctaataattg ttctgataat tatttttagtg atatttaa
12781 ctggacaaaag ctgaactaaa gctctcgcac cagcagcatc ttcaatttca ttaacaataa
12841 ccctattata tctattttaa aagaagtcaa tagcataata accttccctt aggcgtttag
12901 ctattttctt tatttttctt ttagtaaadc actttaattt aaacaaggaa acttcagcac
12961 cttgtgaaaa gttagcttta aattgattag cattagaaat ttttttaata actttaatta
13021 tttttccaaa caaaatataa gcacgatatt caactgtgcc aattgataaa aaaggttgaa
13081 caattcattc tgttgcattt tcaatgttta aatgtttgat ctcgtcagca ctattaacta
13141 aatgtacatc ttttccaccg tgtgaattac gtttcttaac gatgacagga aatgatttga
13201 ttgtttcttt actaagaaga gaagaattga cagttagaaa tctattttgt tttaatcatt
13261 tatatgtttc gtattttatcg tttgctataa aaacaacatt aaaaggatta actaaaaaag
13321 tatttatttg attattgggt tttaaaaaat ctacttttct tgaacgattt aaaatcaatt
13381 cagcttcatt aatttttagct tcgaaattaa tgtctttatt ttcaagtaat aagacttcag
13441 catttagttt tttcttttgt aattccttga ttagacttaa agcaaagtgt tgattttttt
13501 caaaatcatc cttgttgtca taaacaacta atgcttttct gtttttaata ttaatttttc
13561 ccattaatct aaattgcttt taaaagctca attgcaagat tagtatttaa atacattgag
13621 cttcttggtt attgcacatt aggattttact tcacaaaaga tcaatgatct gtcttgatca
13681 aacaaaaaat caataccgca ataaaaaagt tgcattactt tactaatttt aactgctaaa
13741 ttttcttggt ctttattcaa aaaaaagcgt tctgccttgg cccctttatt gagattagaa
13801 cgaaaatcac tattattagt tgtatgtaa gcacctataa ctttattggt cacaacaata
:
579961 atgatcctgc aacattagtt gccattgtag tttttaatac gccgccttta ttatttacia
580021 aagaaatgat catatattta aatgattata atatttcttt aatactaaaa aaatac
//
```

問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

- (1) 遺伝子 dnaN の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : a t g a a a a t a
 アミノ酸 : M K I

- (2) 遺伝子 MG_011 の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : c a c c c t t t t
 アミノ酸 : _____

問4. ヒトの DNA 配列のエントリ (NC_000012) のデータを見て以下の間に答えよ。

- (1) 遺伝子 TPI1 が mRNA に転写される領域の最初と最後の DNA 4 文字を書け

最初: _____ 最後: _____

- (2) 遺伝子 TPI1 を構成するエクソンの数を記せ。 _____ 個

- (3) 遺伝子 TPI1 の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の 3 塩基を書け。

第1領域の開始	_____	終結	_____
_____	_____	_____	_____
第2領域の開始	_____	終結	_____
_____	_____	_____	_____

問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

- (1) 遺伝子 dnaN の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : a t g a a a a t a
 アミノ酸 : M K I

- (2) 遺伝子 MG_011 の蛋白質がコードされている DNA 配列の最初の 9 文字と対応するアミノ酸配列の最初の 3 文字を書け。

DNA : a t g g g a a a a
 アミノ酸 : _____

問4. ヒトの DNA 配列のエントリ (NC_000012) のデータを見て以下の間に答えよ。

- (1) 遺伝子 TPI1 が mRNA に転写される領域の最初と最後の DNA 4 文字を書け

最初: _____ 最後: _____

- (2) 遺伝子 TPI1 を構成するエクソンの数を記せ。 _____ 個

- (3) 遺伝子 TPI1 の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の 3 塩基を書け。

第1領域の開始	_____	終結	_____
_____	_____	_____	_____
第2領域の開始	_____	終結	_____
_____	_____	_____	_____

問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

- (1) 遺伝子 dnaN の蛋白質がコードされている DNA 配列の最初の9文字と対応するアミノ酸配列の最初の3文字を書け。

DNA : a t g a a a a t a
 アミノ酸 : M K I

- (2) 遺伝子 MG_011 の蛋白質がコードされている DNA 配列の最初の9文字と対応するアミノ酸配列の最初の3文字を書け。

DNA : a t g g g a a a a
 アミノ酸 : M G K

問4. ヒトの DNA 配列のエントリ (NC_000012) のデータを見て以下の間に答えよ。

- (1) 遺伝子 TPI1 が mRNA に転写される領域の最初と最後の DNA 4文字を書け

最初: _ _ _ _ 最後: _ _ _ _

- (2) 遺伝子 TPI1 を構成するエクソンの数を記せ。 _____ 個

- (3) 遺伝子 TPI1 の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の3塩基を書け。

第1領域の開始 | _____ 終結 | _____
 _____ | _____ _____ | _____
 第2領域の開始 | _____ 終結 | _____
 _____ | _____ _____ | _____

NCBI Sequence Viewerでの表示(MG_011)

The screenshot shows the NCBI Sequence Viewer interface for the MG_011 gene. The top panel displays the gene models for MG_010, MG_011, and MG_012. The MG_011 gene model is highlighted with a blue dashed circle. A blue arrow points from this circle to a zoomed-in view of the MG_011 CDS region. The zoomed-in view shows the DNA sequence (positive strand) and the corresponding amino acid sequence. The DNA sequence is TTTTTC C C A T T P A T C T A A R A T T G C T T T T. The amino acid sequence is T T A A C G A A A K. The amino acid sequence is shown in a red box, and the corresponding DNA sequence is shown in a green box.

ヒトのDNA配列のエントリーの例:ヘッダ一部

```
LOCUS       NC_000012                3287 bp    DNA        linear     CON 03-MAR-2008
DEFINITION  Homo sapiens chromosome 12, reference assembly, complete sequence.
ACCESSION   NC_000012 REGION: 6846967..6850253
VERSION     NC_000012.10  GI:89161190
DBLINK      Project:168
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
             Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3287)
  AUTHORS   Scherer,S.E., Muzny,D.M., Buhay,C.J., Chen,R., Cree,A., Ding,Y.,
             Dugan-Rocha,S., Gill,R., Gunaratne,P., Harris,R.A., Hawes,A.C.,
             :
             Kucherlapati,R., Weinstock,G. and Gibbs,R.A.
  CONSRTM   Baylor College of Medicine Human Genome Sequencing Center Sequence
             Production Team
  TITLE     The finished DNA sequence of human chromosome 12
  JOURNAL   Nature 440 (7082), 346-351 (2006)
  PUBMED   16541075
REFERENCE   2  (bases 1 to 3287)
  AUTHORS   International Human Genome Sequencing Consortium.
  TITLE     Finishing the euchromatic sequence of the human genome
  JOURNAL   Nature 431 (7011), 931-945 (2004)
  PUBMED   15496913
COMMENT     GENOME ANNOTATION REFSEQ:  Features on this sequence have been
```

ヒトのDNA配列のエントリーの例:FEATURE表

```
FEATURES             Location/Qualifiers
     source            1..3287
                       /organism="Homo sapiens"
                       /mol_type="genomic DNA"
                       /chromosome="12"
     gene              1..3287
                       /gene="TPI1"
                       /note="Derived by automated computational analysis using
                       gene prediction method: BestRefseq. Supporting evidence
                       includes similarity to: 1 mRNA"
     mRNA              join(1..140,1323..1446,1558..1642,1717..1849,2147..2232,
                       2508..2595,2724..3287)
                       /gene="TPI1"
                       /product="triosephosphate isomerase 1"
                       /exception="unclassified transcription discrepancy"
                       /note="Derived by automated computational analysis using
                       gene prediction method: BestRefseq. Supporting evidence
                       includes similarity to: 1 mRNA"
                       /transcript_id="NM_000365.4"
     CDS               join(26..140,1323..1446,1558..1642,1717..1849,2147..2232,
                       2508..2595,2724..2842)
                       /gene="TPI1"
                       /EC_number="5.3.1.1"
                       /product="triosephosphate isomerase 1"
                       /protein_id="NP_000356.1"
```

ヒトのDNA配列のエントリーの例:塩基配列(1)

```
1 ccttcagegc ctcggtcca gcgccatggc gccctccagg aagttcttcg ttgggggaaa
61 ctggaagatg aacgggcgga agcagagtct gggggagctc atcggcactc tgaacgcggc
121 caaggtgccg gccgacaccg gtaagccctc gccgaggagg ggtctggccg ggccggggcc
181 ggggcccggg caggagtggc agcgcctctc cccgaggccc cgaggccccg aggccggtat
241 ccgcgcgac ctgatgcagg gctgtgggac gagggccgct ggggtccggg caggggcctc
301 gcagccgcag ccccgtcggt gcgtcgaggg ggcagggcgg agcacatgat gcccttggg
361 ctatggggca ggtaaggacg ttttgggtct cctggaggaa ggtggccccg gggcgcgac
421 tggggctgtg cccgccaggc gacgggggta ggagcggagc ccgaggctct gcgggagacc
481 gggggaggct gggccgcgtg ggctcccgc tccctgcgcc ctggcctccc gcgctgtcg
541 ccgcccacg tagccccaga ctctcccct tctcgcggc cgtccgctc cccgcccga
601 gctgctcggg ctccctgagc ccagatctg accccttccc ttcggcaacc tgaacgactc
661 ccgcccaca cgaaggacg cgagcccgtg ccaaacaggc tgagcgattt gggagtgagg
721 agccatcta ccgcttccc caacctgga acagcaaagc gcaaggcctc tgagtcagtt
781 aggtctctgc caccacggg caaaggatgc tctcctccat cctccttccc cctccaccg
841 aaatcggaga gccgcgggcc tgatccaaag aggcattccc ttctcgttca tccccagag
901 gcctcaatac aaaccccagg agttggcccc tctccttttg ctacaaatcc ttgccttgca
961 aaggggaggg gaggatggc tattttagaa gggaaagcagg gttgctccct ggagaatgct
1021 gagtctgtga ggtgcctatg ccgagaatag ctcgaggaaa ttggagcccc agctgttaaa
1081 agagcagagg gcaggtgag ggccgtggcc tctcaggggt atctggaagg ctcttcgagt
1141 tgagtgcaga cccagcctgg gctggaaaat ggacaaaagg catcttgctg gggtgaaaag
1201 ggggagagca gaaccaagaa gaagagggtg agggctgggg ggctccaggg cactggttag
1261 gaattgtggg gaatgaaggc tttctttagt ctcatcccc tgtggtacca tcttgcctc
1321 agaggtggtt tgtgctcccc ctactgccta tctgacttc gcccggcaga agctagatcc
1381 caagattgct gtggctgccc agaactgcta caaagtgact aatggggctt ttactgggga
1441 gatcaggtga gatcgaggtg gagaggggtg tgtgggacct ttccctcact ttccctgtg
1501 aggggaaagc cacaggggtg gctccctgct gaaccttggc ttcatctctt cctttagccc
1561 tggcatgatc aaagactgcg gagccacgtg ggtggtcctg gggcactcag agagaaggca
1621 tgtctttggg gagtcagatg aggttagtag ccaagagaga agataaggga tgtctttttc
```

ヒトのDNA配列のエントリーの例:塩基配列(2)

```
1681 caagaaggat gtctaccaa gtctgtttct caacagctga ttgggcagaa agtggcccat
1741 gctctggcag agggactcgg agtaatcgcc tgcatggggg agaagctaga tgaaggggaa
1801 gctggcatca ctgagaagg tgttttcgag cagacaaaagg tcatcgcagg tatctctgga
1861 gaaagggacc tttgagccta tccagggcca cagagactca gagggtaggg tcaggccctg
1921 gagcctgtct tggccccat gctgatccag aaaagggaaa aggggagggg gagtgacaat
1981 ctttgcttgg ggccatgac ttctccagcc ccaaggtaga tgccacctgg aaatccccca
2041 atgtccacta gggggcagta ggccaccgtt ctctgtactc cggagaaacct ggctggagag
2101 ctctttcttg ttcacccttc cctccatctg tatctctgcc ctgcagataa cgtgaaggac
2161 tggagcaagg tegtccctggc ctatgagcct gtgtgggcca ttggtactgg caagactgca
2221 acaccccaac aggtaaccgg gccagggagc cctgccctca tcccagcctg cctcaatagg
2281 tttggacaga cacagcccac atggggcaac ccctattttc aaagacacag agaccttgaa
2341 cccagagaca gtgacttgtc caagggcatc cagtccaggg cctggcttgg atcagagccc
2401 tggtagctct actcagtcag aaaccacact aagtgtccac tgggtgccagt gatttttccc
2461 cttagagagg cagaaaagg cttacttagg ccagcttctt gttctaggcc caggaagtac
2521 acgagaagct ccgaggatgg ctgaagtcca acgtctctga tgcggtggct cagagcacc
2581 gtatcattta tggagggtgag tggctttggt tcccggctga ggtggagtgg gctgaggact
2641 agactgagcc ctcgacatg gaggtgggga tggggcagac tcatcccatt cttgaccaag
2701 cccttgttct gctcccttcc caggctctgt gactggggca acctgcaagg agctggccag
2761 ccagcctgat gtggatggct tcttgtggg tgggtcttcc ctcaagcccg aattcgtgga
2821 catcatcaat gccaaacaat gagccccatc catcttccc acccttctc ccaagccagg
2881 gactaagcag cccagaagcc cagtaactgc ctttcccctg catatgcttc tgatggtgtc
2941 atctgctcct tccgtgggcc tcatccaaac tgtatcttcc tttactgitt atatcttcac
3001 cctgtaatgg ttgggaccag gccaatccct tctccactta ctataatggt tggactaaa
3061 cgtcaccaag gtggcttctc cttggctgag agatggaagg cgtggtggga tttgctcctg
3121 ggttccctag gccctagtga gggcagaaga gaaacctacc tctccccttct tacaccgtga
3181 ggccaagatc ccctcagaag gcaggagtgc tgccctctcc catggtgccc gtgcctctgt
3241 gctgtgtatg tgaaccacc atgtgaggga ataaacctgg cactagg
```

問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

- (1) 遺伝子 dnaN の蛋白質がコードされている DNA 配列の最初の9文字と対応するアミノ酸配列の最初の3文字を書け。

DNA : a t g a a a a t a
 アミノ酸 : M K I

- (2) 遺伝子 MG_011 の蛋白質がコードされている DNA 配列の最初の9文字と対応するアミノ酸配列の最初の3文字を書け。

DNA : a t g g g a a a a
 アミノ酸 : M G K

問4. ヒトの DNA 配列のエントリ (NC_000012) のデータを見て以下の間に答えよ。

- (1) 遺伝子 TPI1 が mRNA に転写される領域の最初と最後の DNA 4文字を書け

最初: _ _ _ _ 最後: _ _ _ _

- (2) 遺伝子 TPI1 を構成するエキソンの数を記せ。 _____ 個

- (3) 遺伝子 TPI1 の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の3塩基を書け。

第1領域の開始 | 終結 |
 _ _ _ | _ _ _ _ _ _ | _ _ _
 第2領域の開始 | 終結 |
 _ _ _ | _ _ _ _ _ _ | _ _ _

ヒトのDNA配列のエントリーの例: FEATURE表

FEATURES	Location/Qualifiers
source	1..3287 /organism="Homo sapiens" /mol_type="genomic DNA" /chromosome="12"
gene	1..3287 /gene="TPI1" /note="Derived by automated computational analysis using gene prediction method: BestRefseq. Supporting evidence includes similarity to: 1 mRNA"
mRNA	join(1..140,1323..1446,1558..1642,1717..1849,2147..2232,2508..2595,2724..3287) /gene="TPI1" /product="triosephosphate isomerase 1" /exception="unclassified transcription discrepancy" /note="Derived by automated computational analysis using gene prediction method: BestRefseq. Supporting evidence includes similarity to: 1 mRNA"
CDS	/transcript_id="NM_000365.4" join(26..140,1323..1446,1558..1642,1717..1849,2147..2232,2508..2595,2724..2842) /gene="TPI1" /EC_number="5.3.1.1" /product="triosephosphate isomerase 1" /protein_id="NP_000356.1"

ヒトのDNA配列のエントリーの例:塩基配列(1)

```
1 ccttcagegc ctcggtcca gcgcatggc gccctccagg aagttcttcg ttgggggaaa
61 ctggaagatg aacgggcgga agcagagtct gggggagctc atcggcactc tgaacgcggc
121 caaggtgccg gccgacaccg gtaagccctc gccgaggagg ggtctggccg ggccggggcc
181 ggggcccggg caggagtggc agcgccctct cccgaggccc cgaggccccg aggccggtat
241 ccgcgcgac ctgatgcagg gctgtgggac gagggccgct ggggtccggg caggggcctc
301 gcagccgcag cccgctcggg gcgtcgaggg ggcagggcgg agcacatgat gcccttggg
361 ctatggggca ggtaaggacg ttttgggtct cctggaggaa ggtggccccg gggcgcgac
421 tggggctgtg cccgccaggc gacgggggta ggagcggagc ccgaggctct gcgggagacc
481 gggggaggct gggccgcgtg ggctcccgc tccctgcgcc ctggcctccc gcgctgtcg
541 ccgcccacg tagccccaga ctctcccct tctcgcggg cgtccggtc cccgcgccg
601 gctgctcggg ctccctgagc ccagatctg accccttccc ttcggcaacc tgaacgactc
661 ccgcccaca cgaaggacg cgagcccgtg ccaaacaggc tgagcgattt gggagtgagg
721 agccatccta ccgcttccc caacctgga acagcaaagc gcaaggcctc tgagtcagtt
781 aggtctctgc caccacggg caaaggatgc tctcctccat cctccttccc cctccaccg
841 aaatcggaga gccgcgggcc tgatccaaag aggcattccc ttctcgttca tccccagag
901 gcctcaatac aaaccccagg agttggcccc tctccttttg ctacaaatcc ttgccttgca
961 aaggggaggg gaggatggc tattttagaa gggaaagcagg gttgctccct ggagaatgct
1021 gagtctgtga ggtgcctatg ccgagaatag ctcgaggaaa ttggagcccc agctgttaaa
1081 agagcagagg gcagggtag ggccgtggcc tctcaggggt atctggaagg ctcttcgagt
1141 tgagtgcaga cccagcctgg gctggaaaat ggacaaaagg catcttgctg gggtgaaaag
1201 ggggagagca gaaccaagaa gaagagggtg agggctgggg ggctccaggg cactggttag
1261 gaattgtggg gaatgaaggc tttctttagt ctcatcccc tgtggtacca tcttgcctc
1321 agaggtggtt tgtgctcccc ctactgccta tctgacttc gcccgccaga agctagatcc
1381 caagattgct gtggctgccc agaactgcta caaagtgact aatggggctt ttactgggga
1441 gatcaggtga gatcgaggtg gagaggggtg tgtgggacct ttccctcact ttccctgtt
1501 aggggaaagc cacaggggtg gctccctgct gaaccttggc ttcatctctt cctttagccc
1561 tggcatgatc aaagactgcg gagccacgtg ggtggtcctg gggcactcag agagaaggca
1621 tgtctttggg gagtcagatg aggttagtag ccaagagaga agataaggga tgtctttttc
```

ヒトのDNA配列のエントリーの例:塩基配列(2)

```
1681 caagaaggat gtctaccaa gtctgtttct caacagctga ttgggcagaa agtggcccat
1741 gctctggcag agggactcgg agtaatcgcc tgcatggggg agaagctaga tgaaggggaa
1801 gctggcatca ctgagaaggg tgttttcgag cagacaaaagg tcatcgcagg tatctctgga
1861 gaaagggacc tttgagccta tccagggcca cagagactca gagggtaggg tcaggccctg
1921 gagcctgtct tggccccat gctgatccag aaaagggaaa aggggagggg gagtgacaat
1981 ctttgcctgg ggccatgac ttctccagcc ccaaggtaga tgccacctgg aaatccccca
2041 atgtccacta gggggcagta ggccaccgtt ctctgtactc cggagaaacct ggctggagag
2101 ctctttcttg ttcacccttc cctccatctg tatctctgcc ctgcagataa cgtgaaggac
2161 tggagcaagg tegtccctgg ctatgagcct gtgtgggcca ttggtactgg caagactgca
2221 acaccccaac aggtaaccgg gcccaggagc cctgccctca tcccagcctg cctcaatagg
2281 tttggacaga cacagcccac atggggcaac ccctattttc aaagacacag agaccttgaa
2341 cccagagaca gtgacttgtc caagggcatc cagtccaggg cctggcttgg atcagagccc
2401 tggtagctct actcagtcag aaaccacact aagtgtccac tgggtgccagt gatttttccc
2461 cttagagagg cagaaaaggc ctacttagg ccagcttctt gttctaggcc caggaagtac
2521 acgagaagct ccgaggatgg ctgaagtcca acgtctctga tgcgggtggc cagagcacc
2581 gtatcattta tggaggtagg tggctttggt tcccggctga ggtggagtgg gctgaggact
2641 agactgagcc ctcgacatg gaggtgggga tggggcagac tcatcccatt cttgaccaag
2701 cccttgttct gctcccttcc caggctctgt gactggggca acctgcaagg agctggccag
2761 ccagcctgat gtggatggct tcttgtggg ttgtgcttcc ctcaagcccg aattcgtgga
2821 catcatcaat gccaaacaat gagccccatc catcttccc acccttctctg ccaagccagg
2881 gactaagcag cccagaagcc cagtaactgc ctttcccctg catatgcttc tgatggtgtc
2941 atctgtcct tccgtgggcc tcatccaaac tgtatcttcc tttactgttt atatcttcac
3001 cctgtaatgg ttgggaccag gccaatccct tctccactta ctataatggt tggactaaa
3061 cgtcaccaag gtggcttctc cttggctgag agatggaagg cgtggtggga tttgctcctg
3121 ggttccctag gccctagtga gggcagaaga gaaacctacc tctcccctct tacaccgtga
3181 ggccaagatc ccctcagaag gcaggagtgc tgcctctcc catggtgccc gtgcctctgt
3241 gctgtgtatg tgaaccacc atgtgaggga ataaacctgg cactagg
```


問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

- (1) 遺伝子 dnaN の蛋白質がコードされている DNA 配列の最初の9文字と対応するアミノ酸配列の最初の3文字を書け。

DNA : a t g a a a a t a
アミノ酸 : M K I

- (2) 遺伝子 MG_011 の蛋白質がコードされている DNA 配列の最初の9文字と対応するアミノ酸配列の最初の3文字を書け。

c a c c c t t t t
DNA : a t g g g a a a a
アミノ酸 : M G K

問4. ヒトの DNA 配列のエントリー (NC_000012) のデータを見て以下の間に答えよ。

- (1) 遺伝子 TPI1 が mRNA に転写される領域の最初と最後の DNA 4文字を書け

最初: c c t t 最後: t a g g

- (2) 遺伝子 TPI1 を構成するエクソンの数を記せ。 7 個

- (3) 遺伝子 TPI1 の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の3塩基を書け。

第1領域の開始 | 終結 |
 — — — | — — — — — — | — — —
第2領域の開始 | 終結 |
 — — — | — — — — — — | — — —

ヒトのDNA配列のエントリーの例: FEATURE表

FEATURES	Location/Qualifiers
source	1..3287 /organism="Homo sapiens" /mol_type="genomic DNA" /chromosome="12"
gene	1..3287 /gene="TPI1" /note="Derived by automated computational analysis using gene prediction method: BestRefseq. Supporting evidence includes similarity to: 1 mRNA"
mRNA	join(1..140,1323..1446,1558..1642,1717..1849,2147..2232,2508..2595,2724..3287) /gene="TPI1" /product="triosephosphate isomerase 1" /exception="unclassified transcription discrepancy" /note="Derived by automated computational analysis using gene prediction method: BestRefseq. Supporting evidence includes similarity to: 1 mRNA"
CDS	/transcript_id="NM_000365.4" join(26..140,1323..1446,1558..1642,1717..1849,2147..2232,2508..2595,2724..2842) /gene="TPI1" /EC_number="5.3.1.1" /product="triosephosphate isomerase 1" /protein_id="NP_000356.1"

ヒトのDNA配列のエントリーの例：塩基配列(1)

```

1 ccttcagegc ctcggetcca gcgccatggc gccctccagg aagttcttcg ttgggggaaa
61 ctggaagatg aacgggcgga agcagagtct gggggagctc atcggcactc tgaacgcggc
121 caaggtgccg gccgacaccg gtaagccctc gccgaggagg ggtctggccg ggccggggcc
181 ggggcgggg caggagtggc agcgccctct cccgaggccc cgaggccccg aggccggtat
241 ccgcgcggac ctgatgcagg gctgtgggac gagggccgct ggggtccggg caggggcctc
301 gcagccgag ccccgtcggt gcgctgaggg ggcagggcgg agcacatgat gcccccttga
361 ctatggggca ggttaaggacg ttttgggtct cctggaggaa ggtggccccc gggcgcgcac
421 tggggctgtg cccgccaggc gacgggggta ggagcggagc ccgaggctct gcgggagacc
481 gggggagget gggccgcgtg ggctccccgc tccctgcgcc ctggcctccc gcgctgtg
541 ccgcgcacg tagcccaga ctccctccct tctcgcggg cgtccgctc cccgcgccga
601 gctgctcggg ctccctgagc cccagatctg accccttccc ttcggaacc tgaacgactc
661 ccgcttcca cggaaggac cgagcccgtg ccaaacaggc tgagcgattt gggagtggc
721 agccatccta ccgcttccc caacctgaa acagcaaagc gcaaggcctc tgagtcagtt
781 aggtctctgc ccccacggg caaaggatg tctcctccat cctcctccc cctccaccg
841 aaatcggaga cccgcgggcc tgatccaaag aggcattccc ttctcgttca tccccagag
901 gcctcaatac aaaccacagg agttggcccc tctcctttg ctacaaatcc ttgccttgca
961 aagggagggt gaggatggc tattttagaa gggaaagcagg gttgctccct ggagaatgct
1021 gagtctgtga ggtgcctatg ccgagaatag ctcgaggaaa ttggagcccc agctgttaaa
1081 agagcagagg gcagggtag ggccgtggcc tctcaggggt atctggaagg ctcttcgagt
1141 tgagtgcaga cccagcctgg gctggaaaat ggacaaaagt catcttgctg gggtgaaaag
1201 ggggagagca gaaccaagaa gaagagggtg agggctgggg ggctccaggg cactggtttag
1261 gaattgtggg gaatgaaggc tttctttagt ctcatcccc tgtggtacca tcttgcctc
1321 agaggtggtt tgtgctcccc ctactgccta tatcgacttc gcccggcaga agctagatcc
1381 caagattgct gtggctggc agaactgcta caaagtactc aatggggctt ttactgggga
1441 gatcaggtga gatcgagggt gagaggggtg tgtgggacct tcccctcact tccctggtg
1501 aggggaaagc cacaggggtg gtcctctgct gaaccttggc ttcattctt cctttagccc
1561 tggcatgac aaagactgag gagccacgtg ggtggtcctg gggcactcag agagaaggca
1621 tgtctttggg gagtcagatg aggttagtag ccaagagaga agataaggga tgtctttttc
    
```

問3. マイコプラズマ菌の全ゲノム配列データ (NC_000908) のデータを見て、以下の問いに答えよ。

(1) 遺伝子dnaNの蛋白質がコードされているDNA配列の最初の9文字と対応するアミノ酸配列の最初の3文字を書け。

DNA : a t g a a a a t a
 アミノ酸 : M K I

(2) 遺伝子MG_011の蛋白質がコードされているDNA配列の最初の9文字と対応するアミノ酸配列の最初の3文字を書け。

DNA : a t g g g a a a a
 アミノ酸 : M G K

問4. ヒトのDNA配列のエントリー (NC_000012) のデータを見て以下の間に答えよ。

(1) 遺伝子TPI1がmRNAに転写される領域の最初と最後のDNA4文字を書け
 最初: c c t t 最後: t a g g

(2) 遺伝子TPI1を構成するエキソンの数を記せ。 7 個

(3) 遺伝子TPI1の蛋白質がコードされた領域 (CDS) のうち第一、第二の領域の開始、終結点前後の3塩基を書け。

第1領域の開始 | 終結 |
g c c | a t g c c g | g t a
 第2領域の開始 | 終結 |
c a g | a g g c a g | g t g

NCBI Sequence Viewerでの表示(TPI1)

エキソンが7つ

第一領域

第二領域

第一領域

第二領域

RefSeq

NCBIで管理されている参照配列データベース
(RefSeqはReference Sequenceの略)

GenBankのデータは、研究者がそれぞれ登録している
ので、塩基配列の精度、アノテーションの質・量
にばらつきが大きい。また、重複しているデータも多い。

RefSeqでは、いくつかのモデル生物に絞り、
それらを専門家が整理して、重複がなく、かつ一定の質の
アノテーションになるように、再構成したデータベース

※アクセッション番号の最初の文字が、
染色体はNC, mRNAはNM、タンパク質はNPとなっている。

蛋白質の配列のデータベース

nr

- ・GenBankに登録されている核酸データのFEATUREに記載されているアミノ酸配列を集めたデータベース
- ・配列数は1000万ぐらい(2010年4月)
- ・蛋白質の名前やその機能については、統一した記載がされていない
- ・配列の精度が低く、同じような配列が少しずつ異なる長さで登録されていることもある。

蛋白質の配列のデータベース

Uniprot

- ・GenBank等のアミノ酸配列データをもとに、専門職員が一つずつチェックし、冗長性を除き、統一した形式で、命名・機能情報を記載したデータベース
- ・配列数は51万ぐらい(2010年3月)
- ・以前はSWISS-PROTと呼ばれていたが、他のデータベースを統合してUniprotの名称となった。
- ・[遺伝子名]_[生物種名]の形式のID。
例) HBA_HUMAN, HBA_MOUSE, TPIS_ECOLI, TPIS_YEAST

Uniprotのデータの例(1)

```
ID TPIS_CHICK Reviewed; 248 AA.
AC P00940;
DT 21-JUL-1986, integrated into UniProtKB/Swiss-Prot.
DT 23-JAN-2007, sequence version 2.
DT 20-JAN-2009, entry version 79.
DE RecName: Full=Triosephosphate isomerase;
DE Short=TIM;
DE EC=5.3.1.1;
DE AltName: Full=Triose-phosphate isomerase;
GN Name=TPI1;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=85166263; PubMed=3885220;
RA Straus D., Gilbert W.;
RT "Chicken triosephosphate isomerase complements an Escherichia coli
RT deficiency.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2014-2018(1985).
RN [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX MEDLINE=86310829; PubMed=3837846;
RA Straus D., Gilbert W.;
```

Uniprotのデータの例(2)

```
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS), AND SEQUENCE REVISION.
RX MEDLINE=75175220; PubMed=1134550; DOI=10.1038/255609a0;
RA Banner D.W., Bloomer A.C., Petsko G.A., Phillips D.C., Pogson C.I.,
RA Wilson I.A., Corran P.H., Furth A.J., Milman J.D., Offord R.E.,
RA Priddle J.D., Waley S.G.;
RT "Structure of chicken muscle triose phosphate isomerase determined
RT crystallographically at 2.5-A resolution using amino acid sequence
RT data.";
RL Nature 255:609-614(1975).
CC !- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone
CC phosphate.
CC !- PATHWAY: Carbohydrate biosynthesis; gluconeogenesis.
CC !- PATHWAY: Carbohydrate degradation; glycolysis; D-glyceraldehyde
CC 3-
CC phosphate from glycerone phosphate: step 1/1.
CC !- SUBUNIT: Homodimer.
CC !- SIMILARITY: Belongs to the triosephosphate isomerase family.
DR EMBL; M11314; AAA49094.1; -; mRNA.
DR EMBL; M11941; AAA49095.1; -; Genomic_DNA.
DR PIR; A23448; ISCHT.
DR RefSeq; NP_990782.1; -.
DR UniGene; Gga.4148; -.
DR PDB; 1SPQ; X-ray; 2.16 A; A/B=1-248.
DR PDB; 1SQ7; X-ray; 2.85 A; A/B=1-248.
DR PDBsum; 1SPQ; -
```

Uniprotのデータの例(3)

```
DR GO; GO:0004807; F:triose-phosphate isomerase activity; IEA:InterPro.
DR InterPro; IPR013785; Aldolase_TIM..
DR Gene3D; G3DSA:3.20.20.70; Aldolase_TIM; 1.
DR PANTHER; PTHR21139; Triophos_ismrse; 1.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
DR TIGRFAMs; TIGR00419; tim; 1.
DR PROSITE; PS00171; TIM; 1.
PE 1: Evidence at protein level;
KW 3D-structure; Direct protein sequencing; Fatty acid biosynthesis;
KW Gluconeogenesis; Glycolysis; Isomerase; Lipid synthesis;
FT INIT_MET 1 1 Removed.
FT CHAIN 2 248 Triosephosphate isomerase.
FT /FTid=PRO_0000090121.
FT ACT_SITE 95 95 Electrophile.
FT ACT_SITE 165 165 Proton acceptor.
FT BINDING 13 13 Substrate.
FT MUTAGEN 95 95 H->N: Reduces activity 5000-fold.
FT CONFLICT 17 18 DK -> KR (in Ref. 3; AA sequence).
SQ SEQUENCE 248 AA; 26620 MW; AFCC258E574DE982 CRC64;
MAPRKFFVGG NWKMNGDKKS LGELIHTLNG AKLSADTEVV CGAPSIYLDF ARQKLDKIG
VAAQNCYKVP KGAFTEGEISP AMIKDIGAAW VILGHSERRH VFGESDELIG QKVAHALAEG
LGVIACIGEK LDEREAGITE KVVFEQTKAI ADNVKDWSKV VLAYEPVWAI GTGKTATPQQ
AQEVHEKLRG WLKSHVSDAV AQSTRIIYGG SVTGGNCKEL ASQHDVDGFL VGGASLKPEF
VDIINAKH
//
```

参考文献

- B.Alberts他著 Essential 細胞生物学 原書 第2版 (2005). 南江堂
- T.A.Brown. ゲノム第3版. (2007). メディカル・サイエンス・インターナショナル.
- 加納 圭 ヒトゲノムマップ (2007). 京都大学出版会