微生物ゲノムクイックスタート

このミニコースでは、微生物のゲノム配列とアノテーションにどのようにアクセスし、どのようにデータを見れば よいか、また遺伝子やアミノ酸配列情報をどのようにダウンロードすればよいかを紹介すると共に、NCBIから提供 されている個別ゲノムや比較ゲノム解析ツールについても紹介します。このコースを通して、「水平伝播したと思 われる遺伝子をバクテリアゲノム中から見つける」であるとか「病原菌と近縁種の常在菌との間の違いは何です か?」といった実践的な問題に対応できるようになることが期待されます。

課題1

■ 課題1-1_

NCBIホームページから、"Genomic Biology"⇒"Genome Resource:Microbial"とリンクをたどって微生物ゲノムプ ロジェクト一覧のページへと移ってください。

ゲノム解読が終了した微生物ゲノムは何種類ありますか?

現在解読中の微生物ゲノムは何種類ありますか?その中でアセンブリング中の状態のものは何種ありますか?

"Firmicutes"に限ると何種のゲノムが解読終了していますか?

■ 課題1-2

解読が終了した微生物ゲノムプロジェクト一覧のページに戻ってください。大腸菌株の箇所までスクロールしてく ださい。

大腸菌では何株のゲノムが完全解読されていますか?

0157堺株を選択して、ゲノムプロジェクトの概要のページに移ってください。

■ 課題1-3_

堺株は、ヒトに対して病原性を持っていますか?

病原性株と非病原性株の違いについてゲノム情報的にどのように違うと書いてありますか?

アクセッション番号NC_002695をクリックしてEntrez Genomeの概要ページへと移動してください。

■ 課題1-4_

Entrez Genomeの概要ページからRefSeqのゲノム情報や、遺伝子情報、タンパク質一覧テーブルへアクセスしてみて下さい。

ゲノム上にいくつの遺伝子がアノテーションされていますか?

ゲノム上の位置や、向き、遺伝子名が書かれているタンパク質一覧の詳細ページへ移動して下さい。

構造RNA一覧の詳細ページへ移動して、FASTAファイルを取得してください。

アルギニンに対応したtRNAはいくつありますか?

Entrez Genomeの概要ページへ戻ってください。

■ 課題1-5_

続いて、Entrez Genomeの概要ページから"COG"と書かれたリンクへと移動してください。 COGとは、NCBIが提供している遺伝子の機能分類データベースであり、各ゲノムのタンパク質コード遺伝子にCOG番号を割り当てることで、遺伝子が持つ機能的な情報も合わせて提供しています。

表中の"% in genome"と"% in genus"の違いに注目してください。何が重要か分かりますか?

表より"Intracellular trafficking and secretion"に分類される遺伝子リストをダウンロードしてください。

Entrez Genomeの概要ページへ戻ってください。

■ 課題1-6_

続いて、Entrez Genomeの概要ページから"TaxPlot"と書かれたリンクへと移動してください。 TaxPlotでは現在見 ている生物種の全遺伝子(染色体、プラスミドすべて)を他の二生物種全遺伝子と比較し、その結果を相同性に基 づいてプロットした情報を提供しています。

病原性大腸菌である0157の特徴を調べるために、大腸菌K12株と他のH157株との比較を行ってみてください。

プルダウンメニューより、カテゴリを選択することで結果に対して機能的な分類を行ってみてください。(現在この機能は提供されていません。)

"Translation and Cell envelope biogenesis"や "Cell motility"に分類された遺伝子の分布を比較してみてください。(現在この機能は提供されていません。)

何が分かりましたか

Entrez Genomeの概要ページへ戻ってください。

解答・解説

■ 解答1-1_

<u>微生物ゲノムプロジェクトー覧ページの立ち上げ</u>

NCBIのトップページから画面左にある"Genomic biology"をクリックし

		Google 👝 🔲 🗙
C NOBL Homerage		N R- 6-
4 7 C W	http://www.ncbi.nim.nin.gov/	
こちらのブックマーク バーにブ	ックマークを適加すると簡単にページにアクセスできます。	🗀 その他のブックマーク
5	National Center for Biotechnology	Information
S NCBI	National Library of Medicine National	I Institutes of Health
PubMed All Da	atabases BLAST OMIM Books Tax	Browser Structure
Search All Databas	es 🗸 for 🛛 🕞 🖓	
OTTE MAD		
Alphabetical List	▶ What does NCBI do?	Hot Spots
Resource Guide	Established in 1988 as a national resource for	Clusters of
About NCBI	molecular biology information, NCBI creates	orthologous groups
An introduction to	public databases, conducts research in computational biology, develops software	Coffee Break.
NCBI	tools for analyzing genome data, and	Genes & Disease,
GenBank	disseminates biomedical information - all for	NCBI Handbook
Sequence	the better understanding of molecular	Electronic PCR
and software	disease. More about NCBI	
Literature	usedse. More about tobi	Entrez Home
databases	Primer-BLAST	Entrez Tools
PubMed, OMIM,	The new Primer-BLAST service designs more	
Books, and PubMed Central	effective and gene-specific PCR primers. The tool combines Primer 3 primer design with a	omnibus (GEO)
Molecular	specificity check provided by a specialized	Human genome
databases	BLAST search. For more information, see	resources
Sequences,	BLAST NEWS.	Influenza Vinie
taxonomy		Resource
Genomic biology	PubMed Central	Man Viewer
The human	PubMed Central is an archive of biomedical	a map fromer
genome, whole	and life sciences journals.	▶ dbMHC
related resources	The second se	Mouse genome
Table	 Free full text Over 1 500 000 articles from over 450 	resources
Data mining	journals	My NCBI
-	 Linked to PubMed and fully searchable 	ing noon
Research at	Use of PubMed Central requires no registration	ORF finder
People, projects,	or fee. Access it from any computer with an	Rat genome
and seminars	Internet connection.	resources
Software		Reference
engineering	NCRI Newe	sequence project

画面右上の"Genome Resources"から"Microbial"をクリックして、微生物ゲノムプロジェクト一覧ページへと移動 してください。



2; Car	npiete Microbiel Genom.	10													
4 ->	C & Http://w	www.nct	inimatingov/geromes/te	initia sul										• [3- 8
15607		101072	INNER-SEPPERCEPT											di0.5 📫	07:07-
S	NCBI	ENT	Genon	ne P	ro	ja	et	;				CD4	nectio Info	discovery	
Case	Postant Descent	NUCLE	tide Postein		0 I	Oann		5	huture	Pa	p li et	Taxonomy		CININ	
Searc	an Genome Project		J			0.00		100							12
organ Tools size	iem group: All legend: T TarMap; a estimated, otherwise - 821 Complete M6	P - Pro	etTable; C - COG Table size is calculated based on I Genomes selected:	D - 3-0	neist	bors; 1	L - ER	AST; 5 - C	200 search: (G - GenePi	lot; X - Teo	Plot; <mark>M</mark> - gMap; F	- FT	P; R - Publice	tiona.
GPID	Organium	King	Group	* Sire	00	1chr	#plam	Gentlank	RetSeg	Released	Modified	Center		Tools	
12997	Acayochisria.matka MERCI1017	.0	Oyanobacteria	8.36	47.0	1	9	CP000628.1	NO.009925.1	10/16/07	07/25/08	Washington University (Washild University	re		EB
19259	Acholeolarra Isidaeli Ethelia	ü	Fimilates	1.5	31.9	1		CP000896.1	NC 010760.1	12/07/07	07/30/08	Ensempti Institute for Provision (Internit Medicine, Internit Medicine, Internit Medicine, Internit Institution Social Onwelcommunit of the Russian Fasteration Lenoral	re	DLSGX	E
15753	Acidehillan cristian	8	Abhaproteobacteria	3.97	67.1	1	8	CP000697.1	NO.009484.1	05/11/07	07/26/08	DOE Joint Genome Initibute	1 E	OLSGX	E
1	Acidthistecilla terroorklans ATCC	0	Gammaprotechacteria	э	58.8	1		CP001219.1	NO.011701.1	12/19/08	12/22/08	1008	е	LSGX	MER
10000	Acidthiobacillus terropoidans ATOC 52003	8	Other Electeria	2.9	58.9	1		CP001122.1	NO 011206 1	09/05/08	09/08/08	DOE Joint Genome Institute	IR		ME
15773	Ackkecteria tacterium Elin345	11	Acidobacteria	5.7	58.4	1		CE000350.1	NO.008009.1	05/04/06	07/20/08	DOE Joint Genome	IE	OLSGX	E
10097	Acidothermus cellulohrticus 11B	11	Actinobacteria	2.4	66.9	1		CP000401.1	NO 000578 1	11/09/06	07/28/08	DCE Joint Genome kuthute	1 E	RLSGX	E
15708	Acidosoras averas	0	Betaprotecbacteria	5.4	68.5	1		CP0005121	NC-0087521	01/04/07	07/29/08	DOE Joint Genome	12	DLSGX	ME
15665	Acidovorau ap. JS42	8	Betaproteobacteria	4.54	66.1	1	2	00000539.1	NO.008782.1	01/04/07	07/21/08	COE Joint Genome	18	DLSGX	ME
21111	Acketskacter Internet AlfXX52	0	Gammaproteobacteria	4.11	39.2	1	1	CP001182.1	NO.011586.1	11/17/08	11/17/08	Case Western Seserve Universite morel	e	LSOX	MER
20992	Acinetubacter teumernil AB007-0294	0	Gammaproteobacteria	3.8	39.0	1		CP001172.1	NC 011595 1	11/24/08	12/11/08	D.MV-Buttako-Coll Intered	e	LSGX	MER
17827	Acinetisbacter Issumernii ACR2/	ġ.	Gammaproteobacteria	3.99	38.9	1	2	02000853.1	NC 010611.1	04/15/08	11/12/08	Institute of molecular biology and pathology. CHE, bay (more)	16		MER
	Achietobacter	-	Commenter to should be be		-			-			07/00/00	Market States			HER
17477	baumannii ATOO 17978		Carrenaprocecoacteria	4.02	30.9			PERSONAL PROPERTY.	OPTIMANE I	0000000	01120100	TOW PARTY.	TC	ALSAVY	NO IS IN

このページには、Entrez Genome Projectのトップページから右側の"Prokaryotic Projects"をクリックすること でも移動することが可能です。



ゲノムプロジェクトの一覧ページでは、1行につき1微生物ゲノムプロジェクトが記載されており、ゲノムサイズや GC含量アクセッション番号とそのリンクや、遺伝子リストへのリンクなどが表示されています。

表の一番上の記載からArchaea55ゲノム、Bacteria766ゲノムの合計821ゲノムについて完成配列が出ていることがわかります。

organ	nism group:	All	*				
Tools	legend: T - Tax	(Map; P - Pro	tTable; C – COG Tab	le; D - 3-D	neighbors	;L-	BLAS
* size	is estimated, othe	rwise genome s	ize is calculated based o	n existing sea	quences		
?	821 Complet	te Microbial	Genomes selecte	I: [A] - 55	б, [B] – 7	66	
GPID	Organism	King	Group	Size	GC #chr	≵ pls	m G

また、表右上から"Genomes in Progress"タブをクリックすると

Organism info	Complete genomes	Genomes in progress
	L	

現在解読中のゲノムプロジェクトの一覧ページへと移動します。

St Minubial Canana	Second a little										164 <mark> 5 B </mark>
6 - C +	titu//www.ncbi.nlm.nlh.at		inian Alimi ka siali								F D+ 4
FART-01-0 15-61	7-52-55-8597832-8880-5-5	27042	2427.								C 70807-07
POWERKT-R.LOW	1774 Column College	1000	129.74								I had been a set
S NCBI	ENTREZ	an	ome Proj	act	3		-	PasSat		1.000	information discovery
Search Genome F	Project 👻		(Clear						Ē	
						Orga	nism	info Compl	lete ger	nomes	Genomes in progress
organism group		-				301153	63.92	Logo	nt 🗆 -	wes assert	by: □ - no sequence available.
1401 Mere	otherwise perione size in caso,	visted on	relacted: [A] - 41	e] - 13	60	equence.	ne avai	lability filter: 😒	all, G	WEI assem	bly; O no sequence available.
1 That men	Occasilant	Mine.	Oran	Conta	00	*	00	Accession	fillent	Released	Center
607 Newton work	0708		Ovarobacteria	-	Course	O SER	-	-	-	(and the second	DOE shint Genome but hat
107 Acarvochioria	REMORE 5410		Overobacteria						2		d Crair Verter buttlite
NO Acetohabble	m arsheticum DSM 5501		Firmicutes		-	-		-	2		DOE Joint Genome Institute
15 Acideminoco	cous intestinals RVC-MR95		Firmicutes	-		-		-	-	-	University of Valencia
100 Acidenus brie	er levil	A	Grenarchaeota			1.0			-		University of Copenhagen
125 Acidenicrobia	m ferroonidana DSM 10331		Actinobacteria	-	-		67- 69	-	-	-	DOE Joint Genome Institute
1 Acidobecterik	am capeulatum ATCC 51196	8	Acidobecteria		-	4.15		-	-	-	TIDE
005 Acidobacteria	am capsulatum ATCC 51196	0	Acidobecteria	-	-	-		-	-	-	J. Craig Venter Institute
105 Acidestras.d	statieidi 2WI	.0	Betaproteobacteria	-	-	-	65- 66		1. Terrer	2	DOE Joint Genome Institute
277 Aciduleroturo	dum boonei T469	A	Euryarchaeota	63	-		39.1	NE AREDOCCOCCO	BLAST	07/15/08	J. Graig Venter Institute
000 Acinetobacte	a baumannii	-00	Gammaproteobacteria	-	-	. 3.9		the second second second	-	-	Genescore
995 Acinetutecte	r barnernii A0000	0	Gammaprotectacteria	68	-	-	38.9	NZ ABROODDOOD	BLADI	12/05/08	Infectious Disesse and Genomics Laboratory, Center of Excellence in BioInformatics and Life Sciences, SJ.NV-Buttalo
333 Achetubecte	r baurarnii MDR-2.00	-10	Gammaprotecbacteria			-			-	-	Designs University, Ohing
09 Acinetobacte	r baumannii WEIA0008	.0	Gammaprotecbacteria		-	-	38.9	÷	-	-	Naval Medical Research Center
SE Achietobacte	r baumannii WEL7215	.0	Gammaprotechacteria	-	-	-	38.9	-	-	-	Naval Medical Finnarch Center
ISS Acinetobacte	r baumannii HEL7255		Gammaprotechacteria	-	-	-	38.9	-	-	-	Neval Medical Research Center
M2 Acketobecte	r. ID. ATOC 27244	.0	Gammaprotectacteria	255	-	-	39,4	NET ARTINE CONCORD	BLAST	10/23/08	Barlor College of Medicine
117 Activobacilla	e aleuropheumoniae seconar 1	0	Gammaprotechacteria	-	-	2.19	37	-	-	-	University of Oklahoma
198 Activobacilla atr. 4274	a steuropreumoniae serovar 1	.0	Gammaproteceacteria	140	2132	* 2.29	41.4	NZ AACHIO000000	BLAST	04/21/03	University of Oklahoma
200 Actinobacilla	a aleuropheumoniae aerovar 5	0	Gammaprotechacteria		-	2.18	37	-	-	-	University of Oklahoma
332 Actinotecilu	a pleuropneumoniae serovar 7	8	Gammaproteobacteria	-	-	\$ 2.08	37	-	-	-	University of Oklahoma
510 Actinomyces	neeskesti Müt	.0	Actinobacteria	-	-	3.06	-	-	-	-	DOR
171 Actionness	odoritoliticus ATOO 17982	30	Actinobacteria	4	2159	-	65.4	00000000/AA	BLAST	03/16/07	Washington University (WashU)
429 Activomyces	urgenitals DOM 15434	- 10	Activobacteria	-	-	-	-	-	-	-	BCM
079 Actinospica r	obinise DOM 44927	D	Actinobacteria		-	3	70.8	-	8	-	DOE Joint Genome Institute
705 Actinosymper	ma minum DGM 43827	.0	Actinobacteria	-	-	-	-	-	-	-	DOE Joint Genome Institute
2172 Accreation	ter activom/cetemcomitana	.0	Gammaprotechacteria		-	-		-	-	-	University of Oklahoma

このページの上部を見ると、現在解読中のゲノムはArchaea41ゲノム、Bacteria1360ゲノムの合計1401ゲノムであることがわかります。

e is estimated, otherwise genome size is calculated based on existing sequences 1401 Microbial Genomes In-Progress selected: [A] - 41, [B] - 1360 King Group Conts CDS Organism

また、画面右上のラジオボタンから"wgs assembly"フィルタを選択することで、

Legend: [- wes assembly; [– no sequ	ence available.
sequencing availability filter: 🔘 all;	💿 wgs assembly;	🔘 no sequ	ence available.
			save

現在ドラフト配列が入手できるゲノムのみに絞り込まれます。その数は、 Archaea8ゲノム、Bacteria608ゲノムの 合計616ゲノムになります。

e	is estimated, otherwise genome s	ize is calculated base	ed on existing seq	uences	
	616 Microbial Genomes	In-Progress sel	ected: [A] - :	8, [B] - 608	}
D	Organism	King	Group	Conts	CD

プロジェクト一覧の絞込み

Г

Г

右上のタブから"Complete Genomes"をクリックして、解読完了ゲノムの一覧に戻ってください。



今度は左上の"organism group:"横のプルダウンメニューから"Frimicutes"を選択して、表示されるゲノムを Firmicutesのものに限ってください。

organ	nism group:	All	~	1					
_	Tools legen	All	1	📢; C – COG	Table; D -	- 3-D	neigh	bors; L	- BLAS
size	is estimated, o	All Archaea		d based on ex	isting seque	inces			
?	767 Comp	Crenarchaeota Eurvarchaeota		elected: [A] - 52,	[B] -	715		
GPID	Organ	Nanoarchaeota		Pup	* Size	GC	#chr	Z plsm	GenBa
2997	Acaryochloris MBIC11017	All Bacteria Acidobacteria		acteria	8.36	47.0	1	9	CP0008
9259	Acholeplasma PG-8A	Actinobacteria Aquificae Bacteroidetes/Chlorobi Chlamydiae/Verrucomic Chloroflexi Cyanobacteria	robia	outes	1.5	31.9	1		<u>CP0008</u>
15753	Acidiphilium cr 5	Firmicutes Fusebacteria Planctomycetes	↓ æ	obacteria	3.97	67.1	1	8	<u>CP0006</u>
6689	Acidithiobacilli ferrooxidans A	Proteobacteria Alphaproteobacteria		acteria	2.9	58.9	1		CP0011
	Acidobacteria	hactorium							

164ゲノムがFirmicutesで解読が終了していることが分かります。

orgai	nism group: Firmicutes Tools legend: T - Ta	xMap;	<mark>∨</mark> P − ProtTa	able; <mark>C</mark> –	COG	Table;	D - 3-I	D nei
size	is estimated, otherwise genome	size is	calculated t	based on ex	kisting s	sequenc	ces	
?	164 Complete Microbia	il Gen	omes se	lected:	A] - [0, [B]	- 164	1
GPID	Organism	King	Group	* Size	GC	#chr	‡ plsm	Gen

■ 解答1-2_____

Г

ゲノムプロジェクトの概要ページの立ち上げ

oraganism group:横のプルダウンメニューからAllを選択して解読が終了している全ゲノムを表示してください。

_

orga	nism group:	All	v	1_					
	Tools leger	All	^	🔋 C – COG 1	Table; <mark>D</mark> -	3-D	neigh	bors; L	-
size	is estimated, o 767 Comp	All Archees Crenarchaeota	à	based on exis	ting seque] – 52,	nces [B] -	715		_
GPID	Organ	Nanoarchaeota		up	* Size	GC	#chr	‡ plsm	G
2997	Acaryochloris MBIC11017	All Bacteria Acidobacteria		acteria	8.36	47.0	1	9	<u>OP</u>
9259	<u>Acholeplasma.</u> PG-8A	Actinobacteria Aquificae Bacteroidetes/Chlorobi Chlamydiae/Verrucomicrobia Chloroflexi Cyanobacteria Deinococcus-Thermus		utes	1.5	31.9	1		CP
15753	Acidiphilium cr 5	Firmicutes Fusobacteria Planctomycetes		obacteria	3.97	67.1	1	8	CP
6689	Acidithiobacillu ferrooxidans A	Proteobacteria Alphaproteobacteria	~	acteria	2.9	58.9	1		CP
	Acidobacteria	hacterium							

画面をスクロールして、大腸菌 (Escherichia coli) を見てみましょう。

<u>13480</u>	Erythrobacter litoralis HTCC2594	в	Alphaproteobacteria	3.05	63.1	1	
<u>16235</u>	<u>Escherichia coli 536</u>	в	Gammaproteobacteria	4.9	50.5	1	
33413	Escherichia coli 55989	в	Gammaproteobacteria	5.2	50.7	1	
<u>16718</u>	Escherichia coli APEC 01	в	Gammaproteobacteria	5.51	50.3	1	2
<u>18083</u>	Escherichia coli ATCC 8739	в	Gammaproteobacteria	4.7	50.9	1	
313	<u>Escherichia coli</u> <u>CFT073</u>	в	Gammaproteobacteria	5.2	50.5	1	
<u>13960</u>	<u>Escherichia coli</u> E24377A	в	Gammaproteobacteria	5.27	50.6	1	6
33409	Escherichia coli ED1a	в	Gammaproteobacteria	5.2	50.7	1	
<u>13959</u>	<u>Escherichia coli HS</u>	в	Gammaproteobacteria	4.6	50.8	1	
<u>33373</u>	Escherichia coli IAI1	в	Gammaproteobacteria	4.7	50.8	1	
<u>33411</u>	<u>Escherichia coli IAI39</u>	в	Gammaproteobacteria	5.1	50.6	1	
<u>32571</u>	<u>Escherichia coli</u> 0127:H6 str. E2348/69	в	Gammaproteobacteria	5.1	50.5	1	2
<u>259</u>	Escherichia coli 0157:H7 EDL933	в	Gammaproteobacteria	5.59	50.3	1	1
<u>27739</u>	<u>Escherichia coli</u> 0157:H7 str. E04115	в	Gammaproteobacteria	5.73	50.4	1	2
<u>226</u>	<u>Escherichia coli</u> 0157:H7 str. Sakai	в	Gammaproteobacteria	5.6	50.5	1	2
<u>33375</u>	Escherichia coli S88	в	Gammaproteobacteria	5	50.7	1	
<u>18057</u>	Escherichia coli SE11	в	Gammaproteobacteria	5.17	50.7	1	6
<u>19469</u>	<u>Escherichia coli SMS-</u> <u>3-5</u>	в	Gammaproteobacteria	5.25	50.5	1	4
<u>33415</u>	<u>Escherichia coli</u> <u>UMN026</u>	в	Gammaproteobacteria	5.2	50.7	1	
<u>16259</u>	<u>Escherichia coli UTI89</u>	в	Gammaproteobacteria	5.21	50.6	1	1
<u>20079</u>	Escherichia coli str. K- <u>12 substr. DH10B</u>	в	Gammaproteobacteria	4.7	50.8	1	
225	Escherichia coli str. K- 12 substr. MG1655	в	Gammaproteobacteria	4.6	50.8	1	
<u>16351</u>	<u>Escherichia coli str. K-</u> <u>12 substr. W3110</u>	в	Gammaproteobacteria	4 .65	50.8	1	
<u>33369</u>	<u>Escherichia fergusonii</u> <u>ATCC 35469</u>	В	Gammaproteobacteria	4.6	49.9	1	
10649	Exiguobacterium sibiricum 255-15	в	Firmicutes	3.01	47.7	1	2

図中赤四角で囲んだ23ゲノムが解読されていることがわかります。

この中から青四角で囲んだ0157 堺株のリンクをクリックして、ゲノムプロジェクトの概要のページに移ってみま しょう。

EQUICA Links	Enterohernorrhadic	Escherichia coli							Contraction of	
Bi Resources									replet	
RLAST	Lineage: Bacteria;	Proteobacteria; Ga	mmaproteobacteria;	Enterobacteriales; Er	terobacteriaceae;	Eschenchie;	Eachench	10 con O15	7.147 str	Sai
penome		and statements of the local division of the	C Protection of the				CORNELS of			
TauPict		100 million (1922)					diana .			
ganisin data		the second se	Contraction of the local division of the loc			16	al and			
Genillane		1000. 4	ALC: NO. 1			16	1	1		
Oenomic		COLUMN AND INCOME.	A CONTRACTOR OF			10				
Protein		10000000				316	1	E		
W00							in and	9		
quenting		Eb	ota: CDC				111			
Column Links			NUMBER OF							
Banger	Genome Projects									_
U. Wat	Plasmid genome:	Int Malancian Disland	. Institute of Dischar	minter and Disusialana	d Managemeiran	DOM DAC DU	ALDO			
attent .	Institute	of Biochemistry and	Physiology of Micros	organisms, Molecular Bi	ology, Russia, Push	chino	DOGM			
attair	Escherichie.co	at University of Br	istol, Department of F	Pathology and Microbio	logy, UK					
808	Escherichie co	F 042 at Samper Inst	itute Bo progress]							
• BRS	Eacherichie.co	101-1 at TIGE [As	sembly]	.1						
Colleri	Escherichia co	F SARU at TIGH LAS	sembly]	ete]						
CyberCell	Escherichia co	AFEC OI at love	State University (Con	nplete]						
E. coli GenExp.	Exchanging co	AICC PAPy at US	DOC Joint Genome In	nstitute (JGI-PGF) (Ca	mpletej	Project data reductoriacease, Exchanchia, Exchanchia coli O157 H7 str Salar reductoriacease, Exchanchia, Exchanchia coli O157 H7 str Salar reductoriacease, Exchanchia, Exchanchia coli O157 H7 str Salar				
E col Will	Escherichie co	B str. RELEXE at 1	Iternational E. coli B.	Consortium Dn progres	e]					
ECCE	Exchanishia co	BITI at TOR And	embly]							
ECOC ECORIDEC	Escherichie.co	BZA at TKOE LAsse	mbly]							
• EPD	Escherichie co	UL21/DEJ/at DOE	Joint Genome Institu	ute [In progress]		Teners in the suboratory strain K-12 ⁻⁷ , DMA Res 2001 Feb Teners and in the study of bacterial genetics, be dominant factuations and endpancing visited (PAIA) Teners and and an induced and endpancing visited (PAIA) Teners and an induced and endpancing visited (PAIA) Teners and an induced and endpancing visited (PAIA) Teners and endpancing visited and endpancing visited and endpancing visited the response of the cell. EHEC attaches and efficies to cells in the large dis in either strain, including at least 9 large PAIa in 0157 H7 d adhesine, as well as finitelial gene clusters and inclusively visited (PAIA)				
ERIC	Escherichie.co	EBL21(DE2) at Kore	a Research Institute	of Bioscience and Biot	echnology On prog	[ase				
nome informa	tion:									
Harner	RefSeq	GenBank	Publications	Length (Mbp)	GC content	Proteins	IBAn	TasMap	C00	co
								~	1	
romosome	NC 002695	BADDODDT	#	5.5	50.5%	5253	141			
omosome smid p0157	NC 002595 NC 002128	BAD00007 AB011549	1	5.5 0.092721	\$0.5% 47.6%	5253 85	141			
romosome smid p0157 smid p054K1	NC 002595 NC 002128 NC 002127	8A000007 A8011549 A8011548	1	5.5 0.092721 0.003306	\$0.5% 47.6% 43.4%	5253 85 3	141			
romosome sumid p0157 sumid p054K1 blications:	NC 002128 NC 002128 NC 002127	AB011549 AB011549	1	5.5 0.092721 0.003306	50.5% 47.6% 43.4%	5253 85 3	141			
romosome umid p0157 smid p054K1 blications: • <u>Havanhi T c</u> 283K1111-2	<u>NC_002595</u> <u>NC_002128</u> <u>NC_002127</u> rf.al. "Complete genom 2	BADDODDT ABD11549 ABD11548	4 1 1	5.5 0.092721 0.003306 ichia coli C157.H7 and p	50,5% 47,6% 43,4%	5253 BS 3	141 	K-12", DN	4 Rez 20	001
romosome semid p0157 semid p054K1 blications: • Hexaphi T (283(1)11-0 Excherichia	bic_002695 bic_002128 bic_002127 tf_al_"Complete genom 2 coll 0157117 str. 54	BADDODOT ABD11149 ABD11549 e sequence of enter skal 0157417	d 1 1	5.5 0.092721 0.003306	50,5% 47,6% 43,4%	5253 BS 3	141 	K-12", DN	4 Rez 20	001
romosome smid pOIS7 smid pOSAK1 blications: • Havanbi T r 2008(1):11-3 Eacherichile cherichile co sor constituent	tr. 002895 bit. 002128 bit. 002128 bit. 002127 tr. al. Complete genom 2 cold 0157417 str. Se 6 This organizm was n colormatry. This enterin of the organizm tests in colormatry. This enterin	ADDITER ADDITER abDITER abDITER e sequence of enter skal O157.H7 vamed for its disco to organism is typics timal microflors. E o	d 1 1 ohemorrhagic Escher Werer, Theodore Esci by present in the low	5.5 0.92721 0.003306 ichia.coli 0157.H7 and p herich, and is one of ar intestine of humans	50.5% 47.6% 43.4% enomic comparison the prectier model where it is the do its host, especially	szsa BS 3 with a laborat	id in the i	K-12 ⁻⁷ , DW study of he se present, noe traits	4 Rex 20 t cterial ; but it is Strains (001 only of £
remosome smid p0357 smid p0541 blications: • Hexabi T. 2838(1):11-2 Eacherichia chevichia co miology, and b or constituent i cause urinam lar processes bare likely to	KC.002895 Nr.002128 Nr.002128 Nr.002127 Nr.4, "Complete genom 2 coll 0157417 str. St Characteristic St College Internet of the complete Inter y tract infections, neo- of the complete Internet y tract infections, neo- to the complete Internet y tract infections, neo-	ABOILER ABOILE	1 1 verer, Theodore Eacl by present in the low in capable of caus invary different inte verechromosomal plas v even have integrate	5.5 0.092721 0.003306 ichia coli 0157.H7 and p tericity, and is one of ar intestine of humans retrail diseases in strail diseases, usually and or within bacteric di into the chromosome	50.5% 47.6% 43.4% genomic compariso the premier model where it is the do by attaching to t phages and distin- through bactericp	organisms up organisms up of the sources of the sources when they ac when they ac	141 ony strain ed in the r ve ansend suite vinde i introdució i integratio	K-12 ⁻⁷ , DW istudy of be present, note traits, is toxins the pathogenici n or transp	4 Rex 20 t scterial ; but it is Straine it disrup ty island settion	001 only of £ t ro s (P
remesence smid pOIST amid pOIST bilications: • <u>Hexapb T</u> ., 2030,1211-2 Eacherichia co- miology, and b or constituen i Cause urinar lular processes is are Bally to cherichia col milifer OIST Ho stirre.	tet. 002895 Nr. 002128 Nr. 002128 Nr. 002128 Nr. 002127 Nr. 002128 Nr. 002128	AB000001 AB01149 AB01149 AB01149 e sequence of enter maned for its disco corparism it typical trial meninghit, and sy be encoded on e horizontally and may is associated with H e of EHEC, and refi	f i i cohemorrhagic Escheri Wy present in the low of is capable of cau imany different inte strachromosomal plas imany different integrate amburger disease, wh octs the specific anti	5.5 0.992721 0.003306 ichis coli 0.157.H7 and p herich, and is one of er intestino of humans, ing various diseases in trial diseases, usually mids or within bacteris ich is caused by the co genic markers found or	50.5% 47.6% 43.4% be premier model where it is the do- its host, especially by attaching in the phages and distri- through bactericp entamination of me the surface of the	organisms up on with a laborat organisms up on an facultat when they ac when they ac when they ac whot cell are t DNA segment age or plasmic at products by o cell EHEC at	d in the we anaerob suite vinde i introdució to termed i integratio enterohem taches and	K-12 ⁻ , (M study of bi present, noe traits, gathogenici n or trainsp orrhagic <i>E</i> effaces to	4 Res 20 t cterial p but it is Straine of t disrup ty island cells in t cells in t	001 only of L EC)
remeasure and pO157 and pO157 and pO157 Dilations: • <u>Herepti T.</u> 2000/115-2 Escherichle ce enology, and b sology, and b <i>Escherichle ce</i> the <i>initial constants</i> <i>Caster univer</i> blar processes is are likely to <i>Cherichle ce</i> thiffer C157HG arises are numeri t excels virial area are numeri t escole virial	to the service state of the service state of the service state of the service state	AB000001 AB01149 AB01159 e sequence of enter askal OJ57477 armed for its disco torqarism it typica tirol microflora. E or torelar meninght, and ty be encoded on e horizontally and ma- is associated with H e of EHEC, and refu	t i i cohemorrhagic Escheri werer, Theodore Esci Wy present in the low of is capable of cau imany different inte strachromosomal plas imany different inte strachromosomal plas imany different integrate amburger disease, wh cots the specific anti iom K-12, hundreds of the locus of enterocy.	5.5 0.92721 0.003306 ichis coli 0157.H7 and j berich, and is one of er intestino of humans, ing various diseases in rind diseases, usually midia or within bacteris rind diseases, usually midia or within bacteris distributed diseases in distributed diseases in the disease of the disease (them associated with in the efficientest, numeri di sp0157.	50.5% 47.6% 43.4% be pretrier model where it is the do- its host, especially by attaching to the oblages and distin- through bacterice ontamination of me- the surface of the genomic islands in 4 us toxins and adh	organisms up invant facultar when they as a hot cell are to phot cell are to a cell EMEC at there strain, inv sein, as well a	ad in the to answer while introducing at a finite and subset of the second integration enteroheer taches and subset at semiconer taches and subset of the second introducing at la	K-12", CM study of bi present, noe traits, g tovins the pathogenici n or trainsp orrhagic <i>E</i> effaces to set 9 large one clusters	4 Res 20 t costerial ; but it is Strains of t disrup ty island settion cosif(EHI cells in t PAIs in i and inor	001 only of £ EC) the 015
romosome Ismid polisti Ismid polisti Iblications: • <u>Homehi T.</u> 2008/11/11-4 <i>Escherichile</i> co- sysiology, and b <i>incherichile</i> co- sysiology, and b <i>incherichile</i> co- sysiology, and b <i>incherichile</i> co- sysiology, and b <i>incherichile</i> co- mo constituent <i>in accessed of the solution</i> <i>in accessed of the solution</i> <i>incherichile</i> co- <i>incherichile</i> co- <i>incherichile</i> co- <i>incherichile</i> co-	Concessor St. 002203 St. 002127 St. 002	AB000001 AB01149 AB01149 e sequence of enter what O1574/7 warned for its disco to comparism is typical to comparison in the heritorative and refit is searcited with it is established with it is establi	a i i cohemorrhagic Escheri dy present in the lon of is carable of caus insany different inte strachromosomal plas even have integrate even have integrate amburger disease, wh ects the specific anti amburger disease, wh ects the specific anti the locus of enterocy to found on the plasm 57 HJ7 was isolated in :	5.5 0.092721 0.003306 ichia coli 0157.H7 and j erichtestino disonae if erichtestino disensee in ring variour disensee in time disensees, usually midia or within bacterie di into the chromosterie di into the chromosterie d	50.5% 47.6% 43.4% genomic comparison where it is the do- its host, expecting to the pression of me through bactering through bactering through bactering through bactering through bactering to the surface of th genomic islands in a sus toxins and adh- ai, Japan.	statistical and a second secon	int in the international international international international internation integration and studing at k fimbrial p	K-12 ⁻⁷ , DN study of his se present, noce trains, getoignet nor transp orrhagic E efficies to sest 9 large me clustern	4 Rex 20 t ccterial ; but it is Strains of t disrup ty island suition colf (EHt cells in t PAIs in ; and inor	001 only of Fo s (P EC) 1 the 1 015
romosome Ismid pojski Ismid pojski Islications: • <u>Havanbi T.</u> 2008/11/11-2 <i>Escherichia</i> co- wiology, and b operoactive operoactive operoactive are liably to <i>incherichia</i> co- string. To the co- string. To the co- string. Islication of the co- Islication of th	tr. 002895 bit. 002127 bit. 00	AB000001 AB01149 AB01149 AB01149 e sequence of enter mand for its disco corparies it typica trial meriophit, and sy be encoded on e and meriophit, and sy be encoded on e the of DHEC, and refi triagaing 0157HT for becretion system, encicy genes are ab ai. This strain of 01: Teatures	t cohemorrhagic Escheri verer, Theodore Esci illy present in the low of is capable of caus many different inte verable of caus many different inte verable of caus many different inte integrate and unified were have integrate int	5.5 0.092721 0.003306 ichia coli 0.157.H7 and j erinteetine of humans, ing various diseases in rind usarese, usually mids or within bacteris trial diseases, usually mids or within bacteris di not the dimensionen genic marriers found or them associated with the effacement, rumer id p0.157.	50.5% 47.6% 43.4% enomic comparison the premier model where it is the dot its host, especially by attaching to t by attaching to t by attaching to the surface of th genomic islands in a sue toxins and adh as, Japan.	assa is with a laborat organisms us ninent fiscular when they as a host call are is host call are to fost call ar	int only strain only strain d in the ive ansend suite virule introducing integration enterohem taches and duding at le	K-12 [°] , <i>DN</i> study of hi present, noc traits, g toine the asthogenic n or training orrhagic <i>E</i> offices to set 9 large ine clustern Temper	A Res 20 toterial ; but it is Strains of toterial toterial toterial policition coll (EH coells in t policitis	001 f perse only of E EC) the li 015
romosome simil polisi bilications: • <u>theoretic T</u> . 2008/111-2 <i>Escherichle</i> ce <i>isology</i> , and b <i>isology</i> , and b	tec. 0022935 NC. 002293 NC. 002128 tec. 00218 tec.	AB000001 AB01149 AB01149 AB01149 e sequence of enter med for its disco corparism it typical trial meninght, and sy be encoded on e trial meninght, and sy be encoded on e horizontally and ma- triaguish 0157 H7 fm secretion system, secretion system, at This strain of 011 Testures ment. Endogs	t cohemorrhagic Escheri werer, Theodore Escl by present in the low of is capable of cau imany different inte strachromosomal pils imany different inte imany different i	5.5 0.092721 0.003306 ichis coli 0157.H7 and j berich, and is one of er intestino of humans, ing various diseases in rind diseases, usually mids or within bacteris rind diseases, usually mids or within bacteris fund the disease in disto the chromosome sich is caused by the co genic markers found or other seasociated with j che officiorment, numeri di sp0157. a 1597 outbreak in Sak	50.5% 47.6% 43.4% enomic comparison the premier model its host, especially by attaching and daths through bacteric portamination of me the surface of th genomic islands in 4 sur toxin and adh aj Japan.	assa source of the second second assaure of the second second assaure of the second second second assaure of the second second second assaure of the second second second second assaure of the second	int the r ony strain d in the r ve are send introducir is termed introducir is termed introducir is termed intregratio enterohem taches and chaling at k f mbrial p	K-12.", GN expresent noe traits, pathogenici orrhagic & effaces to entraits effaces to set 9 large ine cluster Temps	4 Rex 20 t contain ty island saftion coll EH PAIs in t PAIs in t and inor esture Ranse	only only of E EC) the I 015 n up

上図のような概要と、様々な情報へのリンクが記されたページが立ち上がります。

■ 解答1-3____

<u>ゲノムプロジェクト概要ページ</u>

この0157:H7 str.Sakaiのゲノムに関する概要ページを読んでいきましょう。一番下の赤四角で囲んだ部分 に"Pathogenic in:Human"と書いてあることから、このゲノム解読株はヒトに対して病原性を有していることがわ かります。

Escherichia coli 0157:H7 str. Sakai 0157:H7	
Escherichia coll This organism was named for its discoverer, Theodore Escherich, and is one of the premier model organisms us the study of bacterial genetics, physiology, and biochemistry. This enteric organism is typically present in the lower intestine of hus where it is the dominant facultative anaerobe present, but it is only one minor constituent of the complete intestinal microflora. Is capable of causing various diseases in its host, especially when they acquire virulence traits. Strains of <i>E. coll</i> can cause us tract infections, neonatal meningitis, and many different intestinal diseases, usually by attaching to the host cell and introducing hat disrupt normal cellular processes. Virulence proteins may be encoded on extrachromosomal plasmids or within bacteriophage (stinct DNA segments termed pathogenicity islands (PAIs). PAIs are likely to have been transferred horizontally and may even integrated into the chromosome through bacteriophage or plasmid integration or transposition.	ised in imans, E col irinary toxins es and n have
Escherichia coli O157:H7. This strain is associated with Hamburger disease, which is caused by the contamination of meat pro by enterohemorrhagic <i>E. coli</i> (EHEC). The identifier O157:H7 refers to the serotype of EHEC, and reflects the specific an markers found on the surface of the cell. EHEC attaches and effaces to cells in the large intestine.	oducts tigenic
here are numerous differences that distinguish O157:H7 from K-12, hundreds of them associated with genomic islands in either including at least 9 large PAIs in O157:H7 that encode virulence factors. A type III secretion system, the locus of enter offacement, numerous toxins and adhesins, as well as fimbrial gene clusters and iron uptake systems are found in these Pathogenicity genes are also found on the plasmid pO157.	strain, rocyte PAIs.
Escherichia coli O157:H7 strain Sakai. This strain of O157:H7 was isolated in a 1997 outbreak in Sakai, Japan.	

Cellular features						Environn	Temperature			
Gram stain	Shape	Arrangement	Endospores	Motility	Salinity	Oxygen Req.	Habitat	Opt. temp.	Range	
-	Rod	Singles, Pairs		Yes		Facultative	Host-associated	37C	Mesophilic	
Pathogenie	Pathogenic in: Human Disease: Hemorrhagic colitis									

また、その上の青く囲んだ領域に0157など病原性株と非病原性である大腸菌K12のゲノム情報的な違いが述べられています。病原性に関連する遺伝子群は、外来性のプラスミド、バクテリオファージなどによりゲノム中に取り込まれ、周辺配列と明らかに異なる特徴を持った"pathogenicity islands (PAIs) (病原性遺伝子クラスタ) "を形成しています。このPAIs内には、III型分泌系遺伝子群、LEEクラスタ (Locus of Enterocyte Effacement)、多くの毒素関連、接着関連因子、鞭毛関連遺伝子、鉄分取り込み遺伝子などが含まれていることが書かれています。

Genome inform:	ation:					
Name	RefSeq	GenBank	Publications	Length (Mbp)	GC content	Proteins
Chromosome	NC 002695	BA000007	4	5.5	50.5%	5253
Plasmid pO157	NC 002128	AB011549	1	0.092721	47.6%	85
Plasmid pOSAK1	NC 002127	AB011548	1	0.003306	43.496	3
Publications:						
 <u>Havashi T</u> laboratory 	<u>et al.</u> "Complete strain K-12.", D	genome seque <i>NA Res</i> , 2001 F	nce of enterohe eb 28;8(1):11-22	morrhagic Escher	ichia coli 0157	7:H7 and ge

"Genome information"に書かれたリンクから染色体ゲノムのRefSeqである、NC_002695をクリックして "Entrez Genome"の概要ページへ移動して下さい。

Genome	🖌 for		Go	Clear	
s Preview/Index	History Clipb	oard Details			
Overview	 Show 	20 💌 Send to	× •		
1 🛪					
me > Bacteria	> Escherichia	coli O157:Hi	7 str. Sakai, complete ge	nome	
er Bactaria: Dest	aabaataria: Gamm	anntachactaria	Enterphanterialse: Enterphant	terinonna: Eacharichia: Each	arishia sali Esshar
57:H7: Escherichia	coli O157:H7 str.	Sakai	chterobacteriales, Enterobact	erraceae, <u>Escherkonia</u> , <u>Esch</u>	enonia coli, escrier
osomes: genome					
ids: <u>pOSAK1</u> , <u>pO1</u>	157				
Genome Info:	Features:	BLAST homologs:	Links:	Review Info:	
Refseq: NC_002695	Genes: 5372	<u>COG</u>	Genome Project	Publications: [4]	
GenBank: BA000007	Protein coding: 5230	TaxMap	Refseq FTP	Refseq Status: Provisional	
Length: 5,498,450 nt	Structural RNAs: 141	TaxPlot	GenBank FTP	Seq.Status: Completed	
GC Content: 50%	Pseudo genes: None	GenePlot	BLAST	Sequencing center: GIRC	
% Coding: 85%	Others: 170	gMap	TraceAssembly	Completed: 2001/10/02	
Topology: circular	Contigs: 1		<u>CDD</u>	Organism Group	
Molecule: DNA			Other genomes for species: 115		
aene Classification b	ased on COG function	al categories	Search gene, GenelD or I	ocus_tag: Fin	d Gene
◀			Zoom		> 54904
Int		ECx0001 ECx	0004		9,910 M
				ECs0008	mogA

■ 解答1-4_

<u>ゲノム概要のページ</u>

表示された"Entrez Genome"概要のページでは、このゲノムに関連した様々な情報が表示されているとともに遺伝 子情報などへとリンクが張られています。

nome > Bacteria: Prote age: Bacteria: Prote 7:H7: Escherichia coli omosomes: genome mids: pOSAK1, pO1	> Escherichia c obacteria: Gammap i 0157:H7 str. Sakai 57	oli O157:H7 st	r. Sakai, complete ger erobacteriales: Enterobacte	iome riaceae; Escherichia; Esc
Genome Info:	Features	BLAST homolozs:	Links:	Review Info:
Refseq: NC_002695	Genes: 5372	道伝子	Genome Project	Publications: [4]
GenBank: BA000007	Protein coding: 5230	タンパク質コ	──│『遺伝子』	Refseq Status: Provisional
Length: 5,498,450 nt	Structural RNAs: 141	構造RNA(j	Seq.Status: Completed	
GC Content: 50%	Pseudo genes: None	GenePlot	BLAST	Sequencing center: GIRC
% Coding: 85%	Others: 170	gMap	TraceAssembly	Completed: 2001/10/02
Topology: circular	Contigs: 1		CDD	Organism Group
Molecule: DNA			Other genomes for species: 115	

この表から、NC_002695上には5,372の遺伝子、(そのうち)5,230のタンパク質コード遺伝子、141個の構造RNAが アノテートされていることがわかります。

青四角で囲まれた"Protein Coding"と書かれたところをクリックして、タンパク質コード遺伝子の一覧表を表示させてみましょう。遺伝子名やそのゲノム上の位置などが記載されたタンパク質コード遺伝子の一覧が表示されることが確認できます。

Escherichia coli 0157:H7 str. Sa	ikai, con	nplete c	enom	e						
Position: from begin to end	, i	ength: fron	0	to	inf	Ref	resh F	Reset		
Length histogram. Click on a barto select length range.										
		_				5291				
5230 protein(s) shown	ゲノム	ΕØ	- K =	5.10	STA format 🔶	Protein Cl	uster 🔶 S	tructure		
遺伝子名	位置		신다		Gi	GenelD	Locus	Locus_tag	COG(s)	Links
thr operon leader peptide	始め:約	終わり	•	27	15829255	913387		ECs0001		
bifunctional aspartokinase thomeserine dehydrogenase l	354	2816	•	820	15829256	913388	thrA	ECs0002	COG0460E, COG0527E	:**
homoserine kinase	2818	3750	•	310	15829257	913390		ECs0003	COG0083E	:**
threonine synthase	3751	5037	•	428	15829258	913393		ECs0004	C000498E	***
hypothetical protein ECs0005	5251	5547	•	98	15829259	913394		ECs0005		
hypothetical protein ECs0006	5700	6476	· 1	258	15829260	913395		ECs0006	C0G30228	•••
putative inner membrane transport protein	6546	7976		476	15829261	913398		ECs0007	C001115E	•••
transaldolase B	8255	9208	•	317	15829262	913400		ECs0008	C0G0176G	***
molybdenum cofactor biosynthesis protein	9323	9910	•	195	15829263	913401	mogA	ECs0009	C000521H	:**
hypothetical protein ECs0010	9945	10511	•	188	15829264	913402	•	ECs0010	COG1584S	•••
hypothetical protein ECs0012	10660	11373		237	15829265	913403		ECs0012	C0047358	***

次に緑四角で囲まれた"Structual RNAs"と書かれたところをクリックして構造RNAの一覧表を表示させてみましょう。タンパク質コード遺伝子の一覧表と同じような形式で構造RNAの一覧が表示されます。

Position: from	begin	to end	to end Length: from 0				to inf		
ength histogram.	oth range.								
	-								
1 RNA(s) shown									
egends:<> DNA regi Product Name	on in flatfil Start	e format 🔶	DNA regi Strand	on in FAS	STA forma	Locus	Locus tag	Links	
6S ribosomal RNA	227102	228643	+	1542	913977	rrsH	ECs5366	*	
le tRNA	228712	228788	+	77	913984	ileV	ECs5367	**	
Na tRNA	228831	228906	+	76	913988	alaV	ECs5368	*	
23S ribosomal RNA	229090	231992	+	2903	913990	rrlH	ECs5369	*	
5S ribosomal RNA	232085	232204	+	120	914016	rnfH	ECs5370	*	
Asp tRNA	232257	232333	+	77	914020	asp∪	ECs5371	* 	
Asp tRNA	240481	240557	+	77	914060	aspV	ECs5372	*	
Thr tRNA	299982	300057	+	76	914369	thrW	ECs5373	* •	
nisc_RNA	542513	542626	+	114	914612	ffs	ECs5378	*	
Arg tRNA	657182	657258	+	77	916955	argU	ECs5379	*	
3In tRNA	776591	776665	· ·	75	917064	ginX	ECs5384	••	
3In tRNA	776703	776777	·	75	917065	gin∨	ECs5385	••	
				77	047000	meth	EC-6206		

一番右のLinksと書かれた箇所のうち黄色のダイヤモンド印(青四角で囲まれた箇所)をクリックして見てください。FASTA形式で配列が表示されます。

1: NC 002695, Reports Escherichia coli[gi:15829254]
>gi 15829254:227102-228643 Escherichia coli 0157:H7 str. Sakai, complete genome
AACAGGAAGAAGCTTGCTTCTTGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAAACTGCCTGAT
GGAGAGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTC
GGGCCTCTTGCCATCGGATGTGCCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGAC
GATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAG
GCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCT
TCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCC
GCAGAAGGAAGGACCGGCTAACTCCGTGCCAGCAGCCGCGGGTAATACGGAGGGTGCAAGCGTTAATCGGAA
I T ACT GGGCGT AAAGCGCACGCCACGGCCGT T TG T AAAG CAAGAT GT GAAAT CCCCCGGCT CAACCT GGGAA
CT GCA TCT CGAT ACT GCAAGCT TGAGT CT CGT AGAGGGGGGGT AGAAT TCCAGGT GT ACCGGT GAAAT GCG
TAGRGATETI GURGURATTA COTO I GUUGAR GUUGUUCUUTI GUGACGAAGACTI GRUGUTI CAGUTI GUGARAGU
GT GGGGGGGGGGGGT TAGGGGT GGT AGT GGGGGGGG
Tranational tion record and random and the concert creating and the concert an
TTACCIGGICI I GACA LOCACAGA ACTILICAGAGA I GGA LIGGI COLTICGGAACIGI GACAGGA GGA I G
CTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCCGCAACGAGCCCAACCGATCCCTTATCC
TTTGTTGCCAGCGGTCCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATG
ACGTCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCGATACAAAGAGAAGCG
ACCTCGCGAGAGCGAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCAT
GAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCTTGTACACACC
GCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAG
TTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCT
TA

構造RNAの一覧表をもう一度見てみましょう。その中でArg tRNAと書かれているRNA遺伝子はいくつあるでしょうか。上の一覧表で赤く囲んだ遺伝子など合計18個あることがわかります。

■ 解答1-5___

<u>COGテーブルの見方</u>

Backボタンを使って"Entrez Genome"概要のページへ戻ってください。

その表の中からCOGと書かれたリンクをクリックしてください。

Genome Info:	Features:	BLAST homologs:	Links:	Review Info: Publications: [4]	
Refseq: <u>NC_002695</u>	Genes: 5372	202	Genome Project		
GenBank: <u>BA000007</u>	Protein coding: 5230	ТахМар	Refseq FTP	Refseq Status: Provisional	
Length: 5,498,450 nt	Structural RNAs: 141	TaxPlot	GenBank FTP	Seq. Status: Completed	
GC Content: 50%	Pseudo genes: None	GenePlot	BLAST	Sequencing center: GIRC	
% Coding: 85%	Others: 170	gMap	TruceAssembly	Completed: 2001/10/02	
Topology: circular	Contigs: 1		CDD	Organism Group	
Molecule: DNA			Other genomes for species: 133		

以下のような0157 Sakai株の各遺伝子にC06番号をアサインしたデータが表示されたページへと移動します。

COGとは、NCBIが提供している遺伝子の機能分類データベースであり、各ゲノムのタンパク質コード遺伝子にCOG番号を割り当てることで、遺伝子が持つ機能的な情報を提供しているものです。同じCOG番号を持つ遺伝子は同じような機能を持つことが期待されます。

COGによる機能分類は約6000種類ほどに分かれており、さらにその上位階層として25種類に分類されています。以下の図では、上部に0157 Sakaiの各遺伝子がどの25種類に分類されるのかを色分けした情報と、下部にその集計情報とが表示されています。



Co	ode COGs		Description	(% in sequence)	(% in genome)	(% in genus)	(% in Gammaproteobacteria)	(% in Bacteria)
	J	194	Translation	3.2247	3.1751	3.3715	4.2982	4.3421
	A	2	RNA processing and modification	0.0332	0.0327	0.0357	0.0311	0.0144
	K	352	Transcription	5.8511	5.8101	6.1126	5.9460	5.9565
	L	311	Replication, recombination and repair	5.1695	5.3682	5.1322	5.6096	4.9420
	в	0	Chromatin structure and dynamics	0.0000	0.0000	0.0000	0.0162	0.0279
	D	36	Cell cycle control, mitosis and meiosis	0.5984	0.6056	0.6661	0.7877	0.7632
	Y	0	Nuclear structure	0.0000	0.0000	0.0000	0.0000	0.0000
	V	53	Defense mechanisms	0.8810	0.9002	0.9607	1.1040	1.2533
	Т	203	Signal transduction mechanisms	3.3743	3.3224	3.4376	4.3089	3.9496
	м	259	Cell wall/membrane biogenesis	4.3052	4.3208	4.5019	4.6689	4.3540
	N	116	Cell motility	1.9282	1.9967	1.9590	2.1616	1.4129
	Z	0	Cytoskeleton	0.0000	0.0000	0.0000	0.0107	0.0120
	W	4	Extracellular structures	0.0665	0.0655	0.0411	0.0241	0.0145
	υ	150	Intracellular trafficking and secretion	2.4934	2.6514	2.5143	2.4173	1.7598

COG分類の概要表を読み取る

COG分類の概要表をよく見てみましょう。

c	ode	COGs	Description	(% in sequence)	(% in genome)	(% in genus)	(% in Gammaproteobacteria)	(% in Bacteria)
0	J	194	Translation	3.2247	3.1751	3.3715	4.2982	4.3421
0	A	2	RNA processing and modification	0.0332	0.0327	0.0357	0.0311	0.0144
0	K	352	Transcription	5.8511	5.8101	6.1126	5.9460	5.9565
	L	311	Replication, recombination and repair	5.1695	5.3682	5.1322	5.6096	4.9420
	В	0	Chromatin structure and dynamics	0.0000	0.0000	0.0000	0.0162	0.0279
D	D	36	Cell cycle control, mitosis and meiosis	0.5984	0.6056	0.6661	0.7877	0.7632
	Y	0	Nuclear structure	0.0000	0.0000	0.0000	0.0000	0.0000
	V	53	Defense mechanisms	0.8810	0.9002	0.9607	1.1040	1.2533
	T	203	Signal transduction mechanisms	3.3743	3.3224	3.4376	4.3089	3.9496
D	M	259	Cell wall/membrane biogenesis	4.3052	4.3208	4.5019	4.6689	4.3540
D	N	116	Cell motility	1.9282	1.9967	1.9590	2.1616	1.4129
	Z	0	Cytoskeleton	0.0000	0.0000	0.0000	0.0107	0.0120
	W	4	Extracellular structures	0.0665	0.0655	0.0411	0.0241	0.0145
	U	150	Intracellular trafficking and secretion	2.4934	2.6514	2.5143	2.4173	1.7598
	0	<u>163</u>	Posttranslational modification, protein turnover, chaperones	2.7094	2.7005	2.7697	3.3036	2.9523
	C	300	Energy production and conversion	4.9867	4.9100	5.3947	4.9092	4.7983
	G	389	Carbohydrate transport and metabolism	6.4661	6.3830	7.4358	5.1274	4.8607
	E	436	Amino acid transport and metabolism	7.2473	7.1358	7.7859	7.5364	7.2550
0	F	<u>91</u>	Nucleotide transport and metabolism	1.5126	1.4894	1.7072	1.7399	1.7834
	H	160	Coenzyme transport and metabolism	2.6596	2.6187	2.8161	3.1682	2.9938
0	I	113	Lipid transport and metabolism	1.8783	1.8494	1.9643	2.5132	2.9319
	P	314	Inorganic ion transport and metabolism	5.2194	5.1391	5.2715	4.9139	4.5646
	Q	87	Secondary metabolites biosynthesis, transport and catabolism	1.4461	1.4403	1.5929	1.9177	2.3300
	R	595	General function prediction only	9.8903	9.7872	9.8645	9.9953	10.4301
0	S	422	Function unknown	7.0146	6.9067	6.4519	6.5640	6.1355
83	-	1266	Not in COGs	21.0439	21.3912	18.2128	16.9267	20.1622
1000								

この表では、左から1文字コード、現在見ているゲノム配列に含まれる遺伝子数、分類の説明、現在見ているゲノム配列中に占める割合、現在見ている生物種ゲノム全体に占める割合、現在見ている生物種が属する"属"全体に占める割合、γプロテオバクテリア中で占める割合、バクテリア中で占める割合を示しています。

例えば、現在見ている生物種ゲノム全体に占める割合と現在見ている生物種が属する"属"全体に占める割合とを比べることで、この生物種が持つ機能的な特徴を浮き上がらせることができます。

0157 Sakaiでは、細胞外構造物に分類される(W)遺伝子の割合が高いことや、糖代謝、輸送に関する遺伝子の割合 が少ないこと(G)などがわかります。

Intracellular trafficking and secretion(U:細胞内輸送、分泌)に分類される遺伝子の一覧を、 Uと書かれた右 横の150をクリックすることで取得してみましょう。

クリックすると下図の様な遺伝子一覧が得られます。

· > C ☆	http://w	ww.ncbi	nlm.nih.gov/	sutilis/c	ogtik.cgi?gi	=176&cog=U	▶ ⊡• .
ちのブックマーク バーに	フックマークを	適加すると	簡単にページにフ	わせんでき	ます。		🗀 その他のブック
scherichia	coli	O157	7:H7 str	. Sal	kai		
Location	strand	Length	PID	Locus	LocusTag	COG	Product
19298.21748	-	816	15829276	=	ECs0022	COG3188NU	putative outer membrane usher protein precursor
21761_22444	-	227	15829277	=	ECs0023	COG3121NU	putative fimbrial chaperone
22494.23027	-	177	15829278	2	ECs0024	COG3539NU	putative fimbrial protein
29618.30112	+	164	15829284	IspA	ECs0030	COG0597MU	lipoprotein signal peptidase
112884_115589	+	901	15829356	secA	ECs0102	COG0653U	preprotein translocase subunit SecA
19011_120213	-	400	15829364	=	ECs0110	COG1459NU	type IV pilin biogenesis protein
120203 121588	-	461	15829365	=	ECs0111	COG2804NU	hypothetical protein ECs0111
121598.122038	-	146	15829366	=	ECs0112	COG4969NU	putative major pilin subunit
157241159841	-	866	15829397	=	ECs0143	COG3188NU	putative outer membrane usher protein
159883.160608	-	241	15829398	=	ECs0144	COG3121NU	putative chaperone protein EcpD
286244.287983	3.000	579	15829510	=	ECs0257	COG1298NU	FhiA
368426.369520	S 3	364	15829604	-	ECs0350	COG3468MU	putative adhesin
379734.383717	+	1327	15829616	=	ECs0362	COG3468MU	AidA-I adhesin-like protein
379734.383717	+	1327	15829616	=	ECs0362	COG5295UW	AidA-I adhesin-like protein
451255.454197	+	980	15829678	=	ECs0424	COG3468MU	putative flagellin structural protein
490916.491248	+	110	15829712	VaiC	ECs0458	COG1862U	preprotein translocase subunit YajC
491276.493123	+	615	15829713	secD	ECs0459	COG0342U	preprotein translocase subunit SecD
493134_494105	+	323	15829714	secE	ECs0460	COG0341U	preprotein translocase subunit SecF
522729.523352	+	207	15829745	cloP	ECs0491	COG0740OU	ATP-dependent Clp protease proteolytic subunit
579895.581250	+	451	15829794	Ξ	ECs0540	COG1538MU	putative outer membrane transport protein
606274.607290	-	338	15829802	=	ECs0548	COG5295UW	adhesin/invasin-like protein
608198.608653		151	15829805	=	ECs0551	COG15850U	hypothetical protein ECs0551
651426.652118	+	230	15829847	=	ECs0593	COG3121NU	putative chaperone
652149.654758	+	869	15829848	1	ECs0594	COG3188NU	putative outer membrane protein
655789.656304	+	171	15829850	=	ECs0596	COG3539NU	putative fimbrial protein
62098.664335	-	745	15829855	nfrB	ECs0601	COG2804NU	bacteriophage N4 adsorption protein B
678260.679642	+	460	15829864	Ξ	ECs0610	COG1538MU	copper/silver efflux system outer membrane protein CusC
740303.740506	+	67	15829919	tatE	ECs0665	COG1826U	twin arginine translocase protein E
825622.826353	-	243	15829996	=	ECs0742	COG3121NU	putative chaperone
826656.828827		723	15829997	=	ECs0743	COG3188NU	truncated outer membrane protein 🕏
859659 860351	+	230	15830026	=	ECs0772	COG0811U	colicin uptake protein TolQ

■ 解答1-6___

<u>TaxPlot情報の見方</u>

Backボタンを使って"Entrez Genome"概要のページへ戻ってください。 その表の中からTaxPlotと書かれたリンクをクリックしてください。

Genome Info:	Features:	BLAST homologs:	Links:	Review Info:
Refseq: NC_002695	Genes: <u>5372</u>	COG	Genome Project	Publications: [4]
GenBank: BA000007	Protein coding: 5230	TaxMap	Refseq FTP	Refseq Status: Provisional
Length: 5,498,450 nt	Structural RNAs: 141	TaxPlot Tax		Seq.Status: Completed
GC Content: 50%	Pseudo genes: None	GenePlot のり	עלע	Sequencing center: GIRC
% Coding: 85%	Others: 170	gMap	TraceAssembly	Completed: 2001/10/02
Topology: circular	Contigs: 1		CDD	Organism Group
Molecule: DNA			Other genomes for species: 133	

以下のような画面へと移動します。



TaxPlotでは現在見ている生物種の全遺伝子(染色体、プラスミドすべて)を他の二生物種の全遺伝子と比較し、 その結果を相同性に基づいてプロットした情報を提供しています。

今見ているページでは、0157 Sakai株の全5,317遺伝子を出芽酵母(横軸)と線虫(縦軸)と比較し、その結果を 相同性に基づいてプロットした結果が表示されています。各点が遺伝子に対応しており、斜め45度の線より右下に あるものは酵母により似ている事を、左上にあるものは線虫とより似ている事を示しています。

病原性大腸菌である0157 sakai株の特徴を調べるために、大腸菌K12株と他のH157株との比較を行いましょう。画 面上部のChoose two species for comparisonのプルダウンメニューからE.coli K12 W3110とE.coli 0157 EDLとを それぞれ選択し、画面下部のCompareボタンをクリックして下さい。

	Oncose two species for companison	
316407	E. coli str. K12 substr. W3110	~
155864	E. coli 0157:H7 str. EC4042	^
100004	E. coli O157:H7 str. EC4045	
	E. coli O157:H7 str. EG4076 E. coli O157:H7 str. EG4113	
Distribu	E. coli 0157:H7 str. E04115	
	E. coli O157:H7 str. EC4196	
	E. coli O157:H7 str. EC4206	
322 hits	E. coli O157:H7 str. EC4401	🏼 jual hits
	E. coli 0157:H7 str. E04486	
	E. coli O157:H7 str. EO508	
	E. coli O157:H7 str. EC869	
	E. coli O157:H7 str. Sakai	
	E. coli O157:H7 str. TW14588	
	E. COILSETT	
	E. coli UTI89	
	E. coli str. K12 substr. DH10B	
	E. coli str. K12 substr. MG1655	
	E. coli str. K12 substr. W3110	×
Zoom	∑ 10000	
216407	E coli etr. K12 substr. W2110	~
010407		×
155864	E. coli O157:H7 EDL933	~
	Erythrobacter	~
Distrib	E. litoralis HTCC2594	
Distrib	ILE. SP. NAMI	
	Escherichia	
322 hits	E. albertii TW07627	upl hits
022 110	E. coli 101-1	quar mua
	E. coli 536	
	E. coli 53038	
	E. coli ATCC 8739	
	E. coli B171	
	E. coli B7A	
	E. coli CFT073	
	E. coli E110019	
	E. coli E24377A	
	E. coli F11	
_	E. coli HS	-
Zoom	E. coli O157:H7 EDL933	×
x20 💙		
τ		
the prot	tein(s) of interest or enter a query string to see the ho	mologs in
Cutoff:	10 Linear scale 🗸 compare	
0		
Query:	compare clear	

すると、下図のように横軸にE. coli 0157 EDL,縦軸にE. coli K12との相同性を示す結果が得られます。



原文更新日: 2005年2月15日 日本語版更新日: 2008年8月23日

All Rights Reserved, Copyright(C) 1997-2008 Japan Science and Technology Agency(JST)

