36 ウイルスタンパク質のKOとオーソログクラスター

○金 昭、古道 美穂、金久 實
京都大学 化学研究所

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　金 昭（京都大学化学研究所）

KEGG Virus

Integrating viruses and cellular organisms

https://www.kegg.jp/kegg/genome/virus.html

Brite hierarchies and tables for viruses

Virus specific Brite hierarchy files and Brite table files are being developed.

Category	Brite file	
Functional classification	03200	Viral proteins (all viral KOs)
	03210	Viral fusion proteins
Taxonomy	08620	KEGG viruses in the NCBI taxonomy (ICTV and Baltimore classifications)
	08621	KEGG viruses in taxonomic ranks (see KEGG TAXONOMY)
Virus-cell interaction	03220	Virus entry
	03222	Virus entry: animal viruses
Disease	08401	Infectious diseases (contains viral infections)
Drug	08307	Antimicrobials (contains antivirals and targets)
Comparison with cellular organsms	01611	RNA polymerase
	01612	DNA polymerase

KEGG birte file 番号

All virus KOs in KEGG

NCBI/ICTV taxonomy with Baltimore classification

KEGG virus taxonomy with fixed taxonomic ranks

Virus entry protein - human receptor interactions

Antiviral drugs and targets

Virus RNA/DNA polymerase

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　金 昭（京都大学化学研究所）

ウイルスタンパク質のオーソログクラスター (vOC)

1. KEGG GENES のウイルス遺伝子 (vg) カテゴリに含まれるタンパク質が対象
2. SSEARCHですべてのvgペア間の配列比較をし、配列長の違いを考慮して類似度を以下のように定義

   corrected identity (ci) = identity * overlap / (overlap + aalen_diff)

3. 各vgごとに ci ≥ 50% のグループを作り、メンバー数が多い順に並べる
4. グループ間の類似度を定義し、この順番で新規または既存のクラスター番号を付与していく
5. グループ間の類似度の定義は改良の余地が多いが、現在は半数以上のメンバーが既存の同じクラスター番号をもつかどうかで判定している

Viral proteins	595,502	
Viruses (tax IDs)	11,487	
Viral proteins with KO	16,142	(2.7%)
KOs	830	
Viral proteins in vOC	474,370	(80%)
vOCs	26,173	

2022年8月24日現在

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　金 昭（京都大学化学研究所）

03200 Viral proteins (all viral KOs)

KEGG **Viral Proteins**

[Brite menu | Copy URL | Help]

Change brite type

- Option
  - One-click mode
  - Row border shading
- Search
  - Go
- ID search
  - Go
- Join
  - Pathway

▼ ▼ ▼

- ▶ **dsDNA-RT viruses**
- ▶ **ssRNA-RT viruses**
- ▶ **dsRNA viruses**
- ▼ **+ssRNA viruses**
  - ▼ Hepatitis A virus (Hepatovirus A)
    - K25922 3C; Hepatitis A virus protease 3C [EC:3.4.22.28]
  - ▼ Hepatitis C virus (Hepacivirus C)
    - K22472 NS5B; HCV RNA-dependent RNA polymerase [EC:2.7.7.48]
    - K22273 NS3; HCV serine protease / helicase NS3 [EC:3.4.21.98 3.6.1.15 3.6.4.13]
    - K22497 NS2; HCV non-structural protein 2 [EC:3.4.22.-]
    - K22271 Core; HCV core protein
    - K24861 E1; HCV envelope glycoprotein E1
    - K22272 E2; HCV envelope glycoprotein E2
    - K19358 HCVgp1; HCV polyprotein
    - K22274 NS4A; HCV non-structural protein 4A
    - K22498 NS4B; HCV non-structural protein 4B
    - K22275 NS5A; HCV non-structural protein 5A
  - ▼ Coronavirus
    - K25025 nsp16; SARS coronavirus 2'-O-methyltransferase [EC:2.1.1.-]
    - K25021 nsp12; SARS coronavirus RNA-directed RNA polymerase [EC:2.7.7.48]
    - K24546 HE; Betacoronavirus hemagglutinin-esterase [EC:3.1.1.53]
    - K25023 nsp14; SARS coronavirus proofreading exoribonuclease [EC:3.1.13.-]
    - K25013 nsp3; SARS coronavirus non-structural protein 3 [EC:3.4.19.12]
    - K25924 nsP3; Alphacoronavirus papain-like proteinase 1/2 [EC:3.4.19.12 3.4.22.-]
    - K25926 nsp5; Alphacoronavirus 3C-like proteinase [EC:3.4.22.-]
    - K25022 nsp13; SARS coronavirus helicase [EC:5.6.2.3 3.6.4.13]
    - K24149 pp1ab; SARS coronavirus replicase polyprotein 1ab
    - K25011 nsp1; SARS coronavirus host translation inhibitor nsp1
    - K25012 nsp2; SARS coronavirus non-structural protein 2
    - K25014 nsp4; SARS coronavirus non-structural protein 4
    - K25015 nsp5; SARS coronavirus 3C-like proteinase
    - K25016 nsp6; SARS coronavirus non-structural protein 6
    - K25017 nsp7; SARS coronavirus non-structural protein 7
    - K25018 nsp8; SARS coronavirus non-structural protein 8
    - K25019 nsp9; SARS coronavirus non-structural protein 9
    - K25020 nsp10; SARS coronavirus non-structural protein 10
    - K25024 nsp15; SARS coronavirus uridylate-specific endoribonuclease
    - K24152 S; SARS coronavirus spike glycoprotein
    - K25001 S1; SARS coronavirus spike protein S1
    - K25002 S2; SARS coronavirus spike protein S2
    - K24324 S; MERS coronavirus spike glycoprotein

Virus KOの例:

K24546 HE; Betacoronavirus hemagglutinin-esterase [EC:3.1.1.53]

KEGG **ORTHOLOGY: K24546** Help

Entry	K24546 KO
Symbol	HE
Name	Betacoronavirus hemagglutinin-esterase [EC:3.1.1.53]
Brite	KEGG Orthology (KO) [BR:ko00001] 09180 Brite Hierarchies 09185 Viral protein families 03200 Viral proteins K24546 HE; Betacoronavirus hemagglutinin-esterase Enzymes [BR:ko01000] 3. Hydrolases 3.1 Acting on ester bonds 3.1.1 Carboxylic-ester hydrolases 3.1.1.53 sialate O-acetylesterase K24546 HE; Betacoronavirus hemagglutinin-esterase Viral proteins [BR:ko03200] +ssRNA viruses Coronavirus K24546 HE; Betacoronavirus hemagglutinin-esterase BRITE hierarchy
Other DBs	GO: 0001681
Genes	VG: 11273168(HE) 11948233(HE) 22807880(HE) 2732801(F) 3200425(HE) 3283263(HE) 39105216(HE) 54124704(HE) 55060278(HE) 921684(HE) Virus taxonomy UniProt
Reference	PMID:15613517 Authors: Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, Wong BH, Poon RW, Cai JJ, Luk WK, Poon LL, Wong SS, Guan Y, Peiris JS, Yuen KY Title: Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. Journal: J Virol 79:884-95 (2005) DOI:10.1128/JVI.79.2.884-895.2005 Sequence: [vg:3200425]
Reference	Authors: Hurdiss DL, Drulyte I, Lang Y, Shamorkina TM, Pronker MF, van Kuppeveld FJM, Snijder J, de Groot RJ Title: Cryo-EM structure of coronavirus-HKU1 haemagglutinin esterase reveals architectural changes arising from prolonged circulation in humans Journal: Nat Commu DOI:10.10
LinkDB	All DBs

機能が文献に報告されたタンパク質を使う

SSDB Paralog Search Result [Ortholog GFIT]

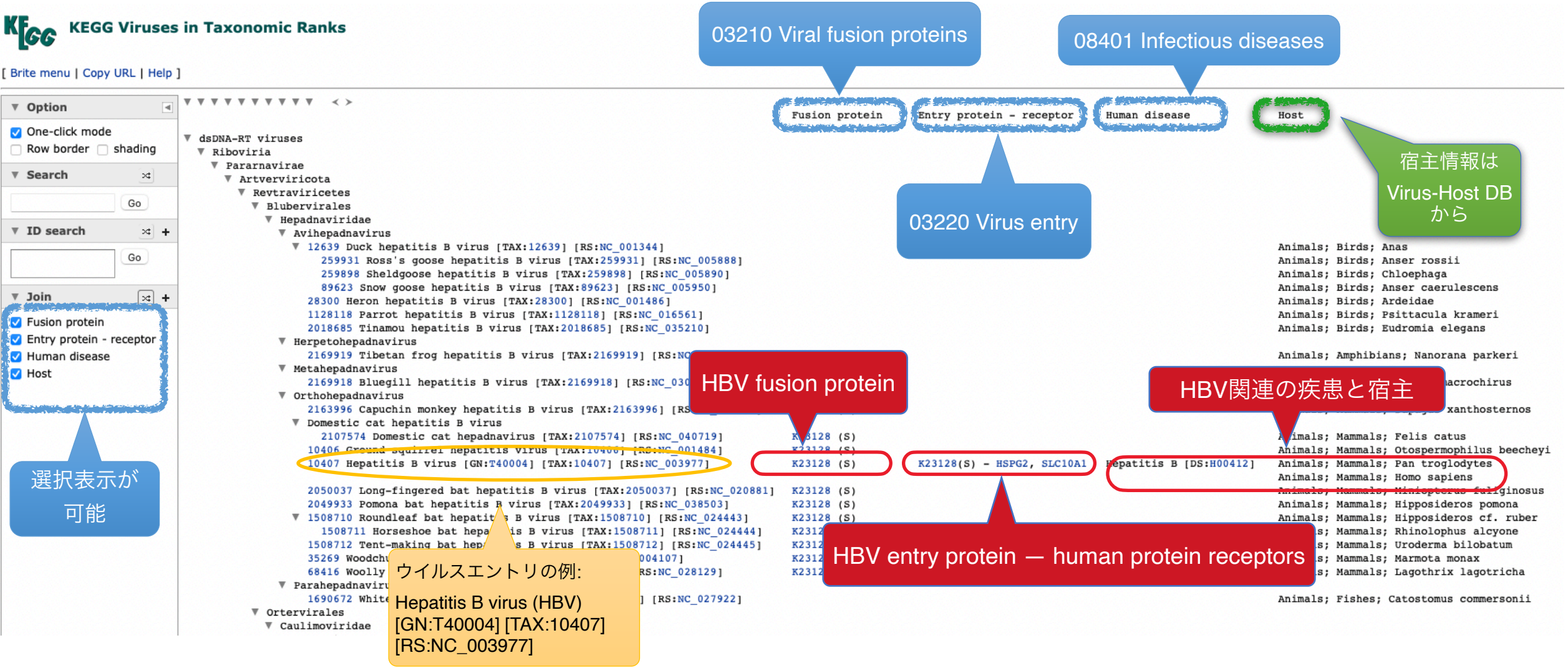
内部用アノテーションツールの例

KO vOC

	SW-Score	Over-lap	Ident	Best	Family	Genus	Definition	KO	vOC
> A vg:3200425 (386)	676	386	1		Coronavi..	Betacoro..	Human coronavirus HKU1; hemagglutinin-esterase glycoprotein	K24546	299
> A vg:11948233 (424)	1511	399	0.561	<->	Coronavi..	Betacoro..	Rabbit coronavirus HKU14; hemagglutinin-esterase protein	K24546	299
> A vg:39105216 (424)	1505	397	0.564	<->	Coronavi..	Betacoro..	Human coronavirus OC43; hemagglutinin-esterase	K24546	299
> A vg:921684 (424)	1497	399	0.561	<->	Coronavi..	Betacoro..	Bovine coronavirus; hemmaglutinin-esterase	K24546	299
> A vg:54124704 (427)	1459	398	0.535	<->	Coronavi..	-	Rodent coronavirus; hemagglutinin-esterase protein	K24546	299
> A vg:11273168 (439)	1443	421	0.506	<->	Coronavi..	Betacoro..	Rat coronavirus Parker; hemagglutinin-esterase	K24546	299
> A vg:22807880 (425)	1436	400	0.527	<->	Coronavi..	Betacoro..	Betacoronavirus HKU24; hemagglutinin-esterase protein	K24546	299
> A vg:3283263 (439)	1417	419	0.501	<->	Coronavi..	Betacoro..	Murine hepatitis virus strain JHM; hemagglutinin-esterase glycoprotein	K24546	299
> A vg:55060278 (312)	878	302	0.467	<->	Coronavi..	Betacoro..	Murine hepatitis virus; hemagglutinin-esterase	K24546	299
> A vg:17603139 (426)	629	378	0.336	<->	Tobanivi..	Torovirus	Porcine torovirus; hemagglutinin esterase		256
> A vg:33133663 (419)	594	421	0.309	<->	Tobanivi..	Torovirus	Goat torovirus; hemagglutinin esterase		256
> A vg:3707767 (419)	591	382	0.33	<->	Tobanivi..	Torovirus	Breda virus; hemagglutinin esterase		256
> A vg:3077359 (655)	415	381	0.297	<->	Orthomyx..	Gammainf..	Influenza C virus (C/Ann Arbor/1/50); hemagglutinin-esterase-fusion	K19253	3115
> A vg:35393478 (664)	395	383	0.303	<->	Orthomyx..	Deltainf..	Influenza D virus (D/swine/Oklahoma/1334/2011); hemagglutinin-esterase precursor	K19253	3115
> A vg:32878304 (667)	364	390	0.277	<->	Alloherp..	Batracho..	Ranid herpesvirus 3; hypothetical protein		
> A vg:32878280 (690)	249	404	0.25	<-	Alloherp..	Batracho..	Ranid herpesvirus 3; hypothetical protein		
> A vg:26628975 (6598)	108	101	0.327	->	Coronavi..	Alphacor..	BtNv-AlphaCoV/SC2013; ORF1ab polyprotein		156
> A vg:18501082 (893)	100	54	0.333	<->	-	Iapetusv..	Erwinia phage PhiEaH1; hypothetical protein		
> A vg:6020270 (6932)	100	86	0.337	->	Coronavi..	Alphacor..	Bat coronavirus 1A; ORF1ab polyprotein		156

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 金 昭（京都大学化学研究所）

08621 KEGG viruses in taxonomic ranks



Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　金 昭（京都大学化学研究所）

03220 Virus entry (文献から抽出されたVirus entry protein - human receptors interaction情報をまとめている)

Virus entry proteinはantiviralドラッグのターゲットになれる?

KEGG Virus Entry

[Brite table menu]

Viral entry protein - human receptor protein interactions

Type	Virus	Taxid	Genome	Entry protein (KO)	Human receptor (hsa ID)	PMID
dsDNA-RT	HBV	10407	T40004	S	HSPG2, (SLC10A1+EGFR)	26776362 26239691 30952782
ssRNA-RT	HTLV-1	11908	T40003	gp46	SLC2A1, NRP1	21994754
	HTLV-2	11909		[illegible]		[illegible]
	HIV-1	11676	T40001	gp120	CD4, CCR5, CXCR4, CD209, CLEC4M	23806647 12960231 1[illegible]21995 1
	HIV-2	11709	T40002	gp120	CD4, CCR5, CXCR4, CLEC4M	12029140 11226297
dsRNA	Rotavirus A	28875	T40092	VP4	(ITGA2+ITGB[illegible]	9144247
				VP7	(ITGAX+ITG[illegible]ITGA4+ITGB1), (ITGAV+ITGB3)	15165605 9144247 11114176
	Rotavirus C	36427	T40094	VP4		[illegible]4247 10590110 12[illegible]
+ssRNA	SARS-CoV-2	2697049	T40361	S		[illegible]07506

HIV-1の例:

gp120 (entry protein) —> CD4, CCR5, CXCR4, CD209, CLEC4M (human receptor proteins)

KEGG Antimicrobials

08307 Antimicrobials

[Brite menu | Copy URL | Help]

Option
- One-click mode
- Row border shading

Search
- Go

ID search
- Go

Join
- ATC
- Disease
- Type
- Target

Target

Antivirals
- Entry, fusion or uncoating inhibitor
  - HIV gp41/gp120
    - [illegible] (USAN/INN) <US> — HIV-1 gp41 [KO:K24261]
    - D10707 Fostemsavir (USAN) — HIV-1 gp120 [KO:K24266]
    - D10708 Fostemsavir tromethamine (USAN) <US> — HIV-1 gp120 [KO:K24266]
    - D10734 Temsavir (USAN) — HIV-1 gp120 [KO:K24266]
  - CCR5 antagonist
    - D03210 Ancriviroc (USAN/INN) — CCR5 (CD195) [HSA:1234] [KO:K04180]
    - D06557 Aplaviroc hydrochloride (USAN) — CCR5 (CD195) [HSA:1234] [KO:K04180]
    - D09878 Cenicriviroc (USAN/INN) — CCR5 (CD195) [HSA:1234] [KO:K04180]; CCR2 (CD192) [HSA:729230] [KO:K04177]
    - D09879 Cenicriviroc mesylate (USAN) — CCR5 (CD195) [HSA:1234] [KO:K04180]; CCR2 (CD192) [HSA:729230] [KO:K04177]
    - D06670 Maraviroc (JAN/INN) <JP/US> — CCR5 (CD195) [HSA:1234] [KO:K04180]
    - D06297 Vicriviroc maleate (USAN) — CCR5 (CD195) [HSA:1234] [KO:K04180]
    - D11399 Leronlimab (USAN/INN) — CCR5 (CD195) [HSA:1234] [KO:K04180]
  - CD4 antibody
    - D09575 Ibalizumab (USAN/INN) <US> — CD4 [HSA:920] [KO:K06454]

IAV	211044 488241 335341 93838 1332244 130760	T40035 T40036 T40037 T40041 T40350 T40[illegible]38	HA	Neu5Acα2-6Gal	G00299 G00255	24161712
IAV	93838 1332244	T40041 T40350	HA	Neu5Acα2-3Gal	[illegible] G00253	24161712
IBV	518987	T40039		Neu5Acα2-6Galβ1-4GlcNAc(6'-SLN)	G00255	9191848

IAV (Influenza A virus)の例:

HA (entry protein) —> Neu5Acα2-6Gal (glycan)

KEGG GLYCAN: G00299

Help

Entry	G00299 Glycan
Composition	(Gal)1 (Neu5Ac)1
Mass	471.4
Structure	Neu5Ac α2——6 Gal G00299 KCF file KCaM SNFG
Other DBs	CCSD: 9422 30062 32176 32196 32250 32251 32455 32841 33254 33911 35021 35246 37405 37948 38321 39411 40555 40575 40746 41572 41692 41807 43995 44009 45348 45712 45743 46201 47052 48476 49112 49195 49272 50818 GlyTouCan: G44345UH JCGGDB: JCGG-STR003577

動物ウイルスentry protein情報は

03222 Virus entry: animal viruses に集約

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 金 昭（京都大学化学研究所）

ウイルスでは実験的データ自体が少ないため、KOとして定義できるものが限られる。そこでKOグルーピングよりも広範囲をカバーするため、vOCを使用してKEGGに含まれるウイルスタンパク質全体の網羅的解析を開始した。

ウイルスにある保存されたタンパク質の例

Virus peptidase

Peptidases and Inhibitors

[Brite menu | Copy URL | Help]

Change brite type

- Option
  - One-click mode
  - Row border shading
- Search
- ID search
- Join
  - Pathway
  - Virus

Virus

- **Aspartic peptidases**
  - Family A2: retropepsin family
    - K22599 PR; HIV protease [EC:3.4.23.16 3.4.23.47] ✓
    - K25054 PR; HTLV-1 protease [EC:3.4.23.-] ✓
- **Cysteine peptidases**
  - Family C1: papain family
    - K20741 VCAT; Autographa californica nucleopolyhedrovirus V-… ✓
  - Family C3: picornain family
    - K25299 3C; Enterovirus protease 3C [EC:3.4.22.28] ✓
    - K25922 3C; Hepatitis A virus protease 3C [EC:3.4.22.28] ✓
    - K25293 2A; Enterovirus protease 2A [EC:3.4.22.29] ✓
  - Family C9
    - K25923 nsP2; Alphavirus protease nsP2 [EC:3.4.22.-] ✓
  - Family C16
    - K25013 nsp3; SARS coronavirus non-structural protein 3 [EC:3… ✓
    - K25924 nsP3; Alphacoronavirus papain-like proteinase 1/2 [EC… ✓
  - Family C27
    - K25925 p150; Rubella virus protease/methyltransferase p150 [… ✓
  - Family C30
    - K25015 nsp5; SARS coronavirus 3C-like proteinase ✓
    - K25926 nsp5; Alphacoronavirus 3C-like proteinase [EC:3.4.22.… ✓
  - Family C37
    - K25325 NS6; Norovirus 3C-like protease [EC:3.4.22.66] ✓
  - Family C57: vaccinia virus I7 peptidase family
    - K25927 I7L; Poxviridae core protease I7 [EC:3.4.22.-] ✓
  - Family C76
    - K21075 UL36, UL48; Herpesviridae large tegument protein dene… ✓

01002 Peptidases and Inhibitors

Virus RNA polymerase

RNA Polymerases

[Brite table menu]

RNA polymerases in prokaryotes and eukaryotes

Bacteria		Archaea		Eukaryotes			Plant		Virus
				Pol II	Pol III	Pol I	Pol IV	Pol V	Poxviridae
K03046(beta') rpoC	K13797 rpoBC	K03041(A') rpoA1 K03042(A") rpoA2		K03006 RPB1 POLR2A	K03018 RPC1 POLR3A	K02999 RPA1 POLR1A	K16250 NRPD1	K16251 NRPE1	K25273 RPO147
K03043(beta) rpoB		K03045(B") rpoB2	K13798 rpoB	K03010 RPB2 POLR2B	K03021 RPC2 POLR3B	K03002 RPA2 POLR1B	K16252 NRPD2/NRPE2		K25274 RPO132
		K03044(B') rpoB1							
K03040(alpha) rpoA		K03047(D) rpoD		K03011 RPB3 POLR2C	K03027 RPAC1 POLR1C				
		K03056(L) rpoL		K03008 RPB11 POLR2J	K03020 RPAC2 POLR1D				
		K03053(H) rpoH		K03013 RPB5/RPABC1 POLR2E					K25277 RPO22
		K03055(K) rpoK		K03014 RPB6/RPABC2 POLR2F					K25279 RPO19
				K03016 RPB8/RPABC3 POLR2H					
		K03059(P) rpoP		K03009 RPB12/RPABC4 POLR2K					
		K03058(N) rpoN		K03007 RPB10/RPABC5 POLR2L					K25278 RPO7

01611 RNA polymerase

Virus DNA polymerase

DNA Polymerases

[Brite table menu]

DNA polymerases in prokaryotes, eukaryotes and viruses

Family	Bacteria		Archaea	Eukaryotes	Viruses
A	K02335(I) polA			K16618(ν) POLN K02349(θ) POLQ	K21314 (T5virus) pol K21238 (T7virus) Gp5
	K02334 dpo			K02332(γ) POLG	
B	K02336(II) polB		K02319 pol	K02324(ε1) POLE K02320(α1) POLA1 K02350(ζ) REV3L K02327(δ1) POLD1	K18942 (T4virus) gp43 K21237 (Phi29virus) gp2 K18963 (Baculovirus) pol K18964 (Herpesvirus) UL30 K21923 (Adenovirus) POL
				K02325(ε2) POLE2 K02321(α2) POLA2	
			K02322 polC		
			K02323 polB	K02328(δ2) POLD2	
C	K02337(IIIα) dnaE K14162 dnaE2	K03763(III) polC			
	K02342(IIIε) dnaQ				K18969 (Bacillus virus CURLY_70

01612 DNA polymerase

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　金 昭（京都大学化学研究所）