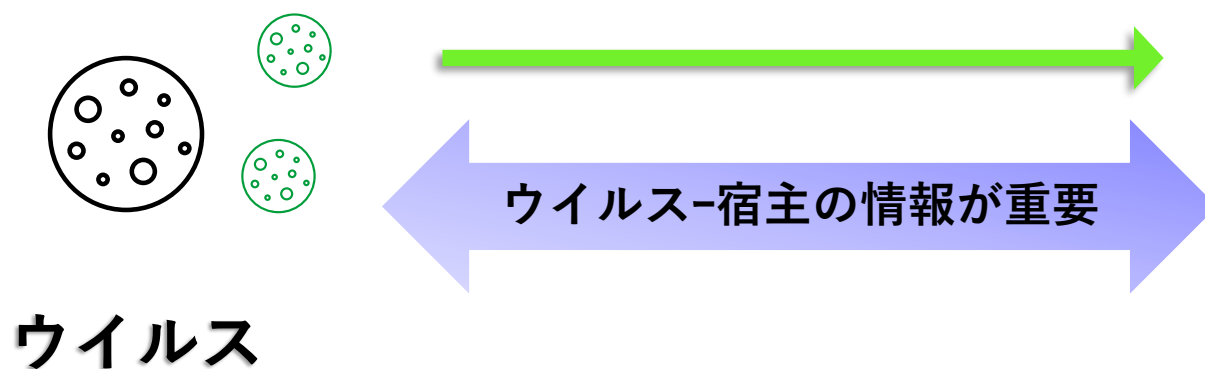


○朱 梦迪、緒方 博之

京都大学化学研究所バイオインフォマティクスセンター

ウイルスは宿主生物を利用して自己を複製する



- ※ ウイルス-宿主間のヌクレオチド・コドン組成の相関
- ※ 遺伝的相互作用
- ※ 共進化・共分化



ウイルス—宿主データベース開発の背景

トーゴの日シンポジウム2021

ウイルス宿主情報へのアクセスは容易ではなかった

Human endogenous retrovirus K113 complete genome

NCBI Reference Sequence: NC_022518.1

FASTA Graphics

Go to:

LOCUS NC_022518 9472 bp DNA linear VRL 30-SEP-2013

DEFINITION Human endogenous retrovirus K113 complete genome.

ACCESSION NC_022518

VERSION NC_022518.1 GI:548558394

DBLINK BioProject: PRJNA222261

KEYWORDS RefSeq.

SOURCE Human endogenous retrovirus K113

ORGANISM Human endogenous retrovirus K113

Viruses; Retro-transcribing viruses; Retroviridae.

REFERENCE 1 (bases 1 to 9472)

AUTHORS Turner,G., Barbarescu,M., Su,M., Jensen-Seaman,M.I., Kidd,K.K. and Lens,J.

TITLE Insertional polymorphisms of full-length endogenous retroviruses in humans

JOURNAL Curr. Biol. 11 (19), 1531-1535 (2001)

PUBMED 11591322

REFERENCE 2 (bases 1 to 9472)

CONSTRM NCBI Genome Project

TITLE Direct Submission

JOURNAL Submitted (29-SEP-2013) National Center for Biotechnology Information, NID, Bethesda, MD 20894, USA

REFERENCE 3 (bases 1 to 9472)

AUTHORS Turner,G., Barbarescu,M., Su,M., Jensen-Seaman,M.I., Kidd,K.K. and Lens,J.

TITLE Direct Submission

JOURNAL Submitted (04-JUN-2001) Molecular Genetics, Albert Einstein College of Medicine, 1360 Morris Park Avenue, Bronx, NY 10461, USA

COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence is identical to AY037928.

COMPLETENESS: full length.

FEATURES

source

Location/Qualifiers

1..9472

/organism="Human endogenous retrovirus K113"

/proviral

/mol_type="genomic DNA"

/host="Homo sapiens"

/db_xref="taxon:106122"

LTR

misc_feature

1112..6746

/standard_name="D22518.1"

gene

6451..8550

/locus_tag="ORF9_01"

宿主はフリーテキスト
大規模処理に不便

Location/Qualifiers

1..9472

/organism="Human endogenous retrovirus K113"

/proviral

/mol_type="genomic DNA"

/host="Homo sapiens"

従来データベースの欠点

- ◆ 標準化
- ◆ 完全性

Refseq+Genbank 78%

NCBI 82%

- ◆ 信頼性

RefSeq の例

Database: RefSeq

Entry: NC_044941

LinkDB: NC_044941

Original site: NC_044941

LOCUS NC_044941 182362 bp DNA linear VRL

DEFINITION African swine fever virus strain L60, complete genome.

ACCESSION NC_044941

VERSION NC_044941.1

DBLINK BioProject: PRJNA485481

KEYWORDS RefSeq.

SOURCE African swine fever virus

ORGANISM African swine fever virus

Viruses; Varidnaviria; Banfordvirae; Nucleocytoviricota; Pokkesviricetes; Asfuvirales; Asfarviridae; Asfivirus.

REFERENCE 1 (bases 1 to 182362)

AUTHORS Portugal,R., Coelho,J., Hoper,D., Little,N.S., Smithson, Upton,C., Martins,C., Leitao,A. and Keil,G.M.

TITLE Related strains of African swine fever virus with different virulence: genome comparison and analysis

JOURNAL J. Gen. Virol. 96 (PT 2), 408-419 (2015)

PUBMED 25408173

REFERENCE 2 (bases 1 to 182362)

CONSTRM NCBI Genome Project

TITLE Direct Submission

JOURNAL Submitted (02-OCT-2019) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

REFERENCE 3 (bases 1 to 182362)

AUTHORS Portugal,R., Coelho,J., Hoper,D., Little,N.S., Smithson, Upton,C., Martins,C., Leitao,A. and Keil,G.M.

TITLE Direct Submission

JOURNAL Submitted (02-AUG-2014) Biochemistry and Microbiology, U of Victoria, 3800 Finnerty Road, Victoria, BC V8P 5C2, Canada

COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to NCBI review. The reference sequence is identical to KM261000.1.

COMPLETENESS: full length.

FEATURES

source

Location/Qualifiers

1..182362

/organism="African swine fever virus"

/mol_type="genomic DNA"

/strain="L60"

/host="domestic pig"

/db_xref="taxon:10497"

/country="Portugal"

/collection_date="1980"

FEATURES

source

Location/Qualifiers

1..182362

/organism="African swine fever virus"

/mol_type="genomic DNA"

/strain="L60"

/host="domestic pig"

/db_xref="taxon:10497"

ウイルス—宿主データベース

トーゴーの日シンポジウム2021

```
ID 001R_FRG3G Reviewed; 256 AA.  
AC Q6GZX4;  
DT 28-JUN-2011, integrated into UniProtKB/Swiss-Prot.  
DT 19-JUL-2004, sequence version 1.  
DT 01-APR-2015, entry version 30.  
DE RecName: Full=Putative transcription factor 001R;  
GN ORFNames=FV3-001R;  
OS Frog virus 3 (isolate Goorha) (FV-3).  
OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Ranavirus.  
OX NCBI_TaxID=654924;  
OH NCBI_TaxID=8295; Ambystoma (mole salamanders).  
OH NCBI_TaxID=30343; Hyla versicolor (chameleon treefrog).  
OH NCBI_TaxID=8404; Lithobates pipiens (Northern leopard frog) (Rana pipiens).  
OH NCBI_TaxID=8316; Notophthalmus viridescens (Eastern newt) (Triturus viridescens).  
OH NCBI_TaxID=45438; Rana sylvatica (Wood frog).  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed=15165820; DOI=10.1016/j.virol.2004.02.019;  
RA Tan W.G., Barkman T.J., Gregory Chinchar V., Essani K.;  
RT "Comparative genomic analyses of frog virus 3, type species of the  
RT genus Ranavirus (family Iridoviridae).";  
RL Virology 323:70-84(2004).  
CC -!- FUNCTION: Transcription activation. {ECO:0000305}.  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AY548484; AAT09660.1; -; Genomic_DNA.  
DR RefSeq; YP_031579.1; NC_005946.1.  
DR ProteinModelPortal; Q6GZX4; -
```

UniProtの場合

TaxIDの記載あり、しかし
宿主の登録率が8%



ウイルス—宿主データベース

トーゴの日シンポジウム2021



Virus-Host DB organizes data about the relationships between viruses and their hosts, represented in the form of pairs of NCBI taxonomy IDs for viruses and their hosts. Virus-Host DB covers viruses with complete genomes stored in 1) NCBI/RefSeq and 2) GenBank whose accession numbers are listed in EBI Genomes. The host information is collected from RefSeq, GenBank (in free text format), UniProt, ViralZone, and manually curated with additional information obtained by literature surveys.

➡ Browse all viruses

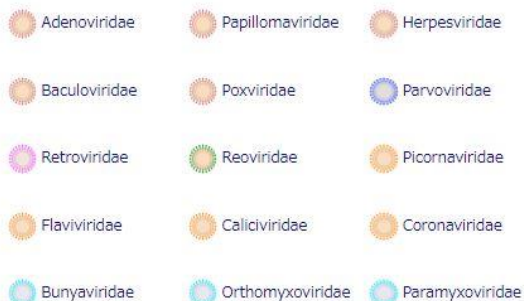


Human viruses

Other major hosts

[Animal | Plant | Other Eukaryotes | Bacteria | Archaea]

Selected viral families



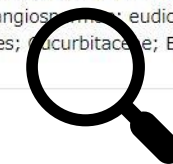
Hop stunt viroid - cucumber

Scientific Name	Hop stunt viroid - cucumber [TAX:12902]
Lineage	Viruses; Pospiviroidae; Hostuviroid; Hop stunt viroid
Baltimore Group	IV/V: ssRNA
Genome Type	Non-segmented
Genbank	X07405 Cucumber pale fruit viroid (CPFV) RNA.
DBLINKS	ViralZone: family

ウイルス情報

Known hosts (2)

Scientific Name	Cucumis sativus [TAX:3659]
Lineage	Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; eudicotyledons; Gunneridae; Pentapetalae; rosids; fabids; Cucurbitales; Cucurbitaceae; Benincaseae; Cucumis
Evidence	Literature
Reference	PMID: 6328420
Authors	Sano T, Uyeda I, Shikata E, Ohno T, Okada Y
Title	Nucleotide sequence of cucumber pale fruit viroid: homology to hop stunt viroid.
Journal	Nucleic Acids Res. 1984 Apr 25;12(8):3427-34.
DBLINKS	KEGG GENOME: T02486



宿主情報

Scientific Name	Solanum lycopersicum [TAX:4081]
Lineage	Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; eudicotyledons; Gunneridae; Pentapetalae; asterids; lamiids; Solanales; Solanaceae; Solanoideae; Solaneae; Solanum; Solanum subgen. Lycopersicon
Evidence	Literature
Reference	PMID: 6328420
Authors	Sano T, Uyeda I, Shikata E, Ohno T, Okada Y
Title	Nucleotide sequence of cucumber pale fruit viroid: homology to hop stunt viroid.
Journal	Nucleic Acids Res. 1984 Apr 25;12(8):3427-34.
DBLINKS	KEGG GENOME: T02665



ウイルス—宿主データベース

トーゴーの日シンポジウム2021

Aeromonas phage 2_D05

Scientific Name	Aeromonas phage 2_D05 [TAX:2588098]
Lineage	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; unclassified Myoviridae
Baltimore Group	I: dsDNA
Genome Type	Non-segmented
RefSeq	NC_054465 Aeromonas phage 2_D05, complete genome.
DBLINKS	KEGG BRITE: NC_054465 ViralZone: family

Baltimore Group
Genome Type

Known hosts (1)

Scientific Name	Aeromonas rivipollensis [TAX:948519]
Lineage	Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales; Aeromonadaceae; Aeromonas
Evidence	RefSeq
Evidence	NCBI Virus
Evidence	Literature
Reference	PMID: 31284428
Authors	Bai M, Cheng YH, Sun XQ, Wang ZY, Wang YX, Cui XL, Xiao W
Title	Nine Novel Phages from a Plateau Lake in Southwest China: Insights into <i>Aeromonas</i> Phage Diversity.
Journal	Viruses. 2019 Jul 5;11(7).
Comment	Strain D05
DBLINKS	KEGG GENOME: T05453

宿主 (TAXID)

文献情報

Strain・Cell lineなどの備考

ウイルス—宿主データベース

トーゴの日シンポジウム2021



Virus-Host DB

[Top](#) | [About](#) | [Statistics](#) | [Virus Index](#) | [Host Index](#) | [Feedback](#)

検索画面

Total: 5 20 viruses per page « Page 1 »

RefSeq ID

Virus (species) name

African swine

Virus lineage

Host name

Host lineage

ウイルスの名前/生物学名

Download

.tsvファイル出力可能

African swine fever virus OURT 88/3 [TAX:443878]	Viruses; Varidnaviria; Bamfordvirae; Nucleocytoviricota; Pokkesviricetes; Asfuvirales; Asfarviridae; Asfivirus; African swine fever virus	Ornithodoros [TAX:6937]	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Chelicerata; Arachnida; Acari; Parasitiformes; Ixodida; Ixodoidea; Argasidae
African swine fever virus Georgia 2007/1 [TAX:874269]	Viruses; Varidnaviria; Bamfordvirae; Nucleocytoviricota; Pokkesviricetes; Asfuvirales; Asfarviridae; Asfivirus; African swine fever virus	Sus scrofa [TAX:9823]	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Artiodactyla; Suina; Suidae; Sus
		Sus scrofa domesticus [TAX:9825]	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Artiodactyla; Suina; Suidae; Sus; Sus scrofa

- Webでブラウズ（検索）できる
(<https://www.genome.jp/virushostdb/>)



操作環境の制限が少ない

- 全データを一括ダウンロードできる
(<https://www.genome.jp/ftp/db/virushostdb/>)



無償一括 Data download

- BLAST検索がGenomeNetで可能
(<https://www.genome.jp/tools/blast/>)



GenomeNet のBLAST検索画面

アクセス方法・特徴・機能

トーゴーの日シンポジウム2021

GenomeNetでのBLAST検索画面

<https://www.genome.jp/tools/blast/>



BLAST Search

BLAST FASTA KEGG2 ③

Enter query sequence: (in one of the three forms)

Sequence ID (Example) mja:MJ_1041

Local file name 選択されていません

Sequence data

Select program and database:

☒ BLASTP (prot query vs prot db)
☐ BLASTX (nucl query vs prot db)

☐ KEGG GENES
☐ Eukaryotes ☐ Prokaryotes ☐ Viruses
☐ Favorite organism code or category

☐ KEGG MGENES
☐ Environmental ☐ Organismal
☐ Favorite samples

☐ Microbial Reference Genes
☐ Ocean (OM-RGC) ☐ Human gut (IGC)

☐ nr-aa (GenBank, UniProt, RefSeq and PDBSTR)
☐ Swiss-Prot ☐ UniProt ☐ RefSeq
☐ PDBSTR

☐ UniRef50 ☐ UniRef90 ☐ UniRef100

☒ Virus-Host Database ②

①

Compute Clear



BLASTP Search Result

Database: `virushostdb.cds.faa`

Protein sequence database entries related to query - 12 hits

Show alignment

Entry			bits	E-val
Top 10	Clear	Select operation	Exec	
☒	vhdba:YP_008719905.1	Rhizoctonia cerealis alphaendornavirus ...	28.9	2.3
☒	vhdba:NP_042773.1	African swine fever virus pB602L Boreoeuth...	28.5	2.3
☒	vhdba:YP_003280846.1	Helicobasidium mompa alphaendornavirus ...	28.5	2.4
☒	vhdba:YP_009702813.1	African swine fever virus BA71V-B602L (...)	28.5	2.5
☒	vhdba:YP_009703532.1	African swine fever virus Benin 97/1 pB...	27.7	4.1
☒	vhdba:YP_009702327.1	African swine fever virus pB602L Sus sc...	27.7	4.2
☒	vhdba:YP_009703331.1	African swine fever virus pB602L Sus sc...	27.7	4.2
☒	vhdba:YP_009703136.1	African swine fever virus B602L Sus scr...	27.7	4.2
☒	vhdba:YP_010086427.1	Spodoptera exempta nucleopolyhedrovirus...	26.9	6.8
☒	vhdba:NP_620728.1	Ustilago maydis virus H1 cap-pol fusion pr...	27.3	7.0
☐	vhdba:YP_009703849.1	African swine fever virus E75 hypotheti...	26.6	8.8
☐	vhdba:YP_009702485.1	African swine fever virus pBA71-B602L B...	26.6	9.8

Show all result

DBGET integrated database retrieval system, GenomeNet



● 他のデータベースに利用（GLOBI、KEGG）

 GENOME: Rotavirus C Help			
Entry	T40094	Viral	Genome
Definition	Rotavirus C		
Taxonomy	TAX:36427		
Lineage	Viruses; Riboviria; Orthornavirae; Duplornaviricota; Resentoviricetes; Reovirales; Reoviridae; Sedoreovirinae; Rotavirus		
Sequence	RS:NC_007543 NC_007544 NC_007545 NC_007546 NC_007547 NC_007569 NC_007570 NC_007571 NC_007572 NC_007573 NC_007574 Taxonomy		
Disease	Rotaviral enteritis [DS:H00975]		
Host	Homo sapiens [GN:hsa]		
Disease	Viral gastroenteritis [DS:H01430]		
Host	Homo sapiens [GN:hsa]		
Other DBs	Virus-HostDB: 36427		
Reference	PMID:10567650		
Authors	James VL, Lambden PR, Deng Y, Caul EO, Clarke IN		
Title	Molecular characterization of human group C rotavirus genes 6, 7 and 9.		
Journal	J Gen Virol 80 (Pt 12):3181-7 (1999) DOI:10.1099/0022-1317-80-12-3181		
Reference	PMID:11864750		
Authors	Chen Z, Lambden PR, Lau J, Caul EO, Clarke IN		



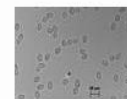
[about](#)
[blog](#)
[browse](#)
[contribute](#)
[data](#)
[datasets](#)
[search](#)
[references](#)
[日本語](#)

Example query: [What do sea otters \(*Enhydra lutris*\) eat?](#) or [What do honey bees \(*Apis*\) pollinate?](#)

What kind of do interacts with according to ?

organisms
interacts with... plenty of things according to Hingamp!

Missing some results? Have suggestions? [Let us know](#). Like a different view? [Open results in interaction browser](#), [the references](#).



Zygosaccharomyces bailii

has pathogen Zygosaccharomyces bailii virus Z



Supported by:

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., and Ogata, H.; Linking virus genomes with host taxonomy. *Viruses* 8, 66 doi:10.3390/v8030066 (2016). Accessed at <ftp://ftp.genome.jp/pub/db/virushostdb/virushostdb.daily.tsv> on 24 Aug 2020. [link show](#) Provider: Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., and Ogata, H.; Linking virus genomes with host taxonomy. *Viruses* 8, 66 doi:10.3390/v8030066 (2016). Accessed via <https://github.com/globalbioticinteractions/virus-host-db/archive/9b2dc140d251480e665709470ae6b55944ba42a7.zip> at 2020-08-24T02:05:45.114Z.

- 2～3 か月ごとに更新
(NCBI/RefSeqのリリースと同期する)
- マニュアル修正・入力



Menu

[Menu | Logout]
Logged in as ogata.
DB Version release201man ^

- Virus-Host Table
- New Accession List
- Removed/Changed Accessions From Previous Release
- Genome Type
 - List of segmented virus
 - List of non-segmented virus
- Genome Composition
- Comparison to NCBI Virus
 - Comparison of Accession
 - RefSeq Accession only in NCBI
 - RefSeq Accession only in VHDB
 - GenBank Accession only in VHDB
 - Comparison of Virus Tax ID
 - Different virus_tax_id between VHDB and NCBI Virus
 - Comparison of Host Tax ID
 - Accession with no host_tax_id in VHDB
 - Accession with no host_tax_id in NCBI Virus
 - Accession with same host_tax_id in VHDB and NCBI Virus
 - Accession with different host_tax_id in VHDB and NCBI Virus
- List of Viruses having both Bacteria/Archaea hosts and Eukaryota host
- Change Log



NC_048666

DB

Virus information

- Virus name: Bacteriophage sp.
- RefSeq Definition: unidentified phage genome assembly, chromosome: 1.
- Lineage: root; Viruses; unclassified viruses; unclassified bacterial viruses; Bacteriophage sp.
- Genome type:
- Other Refseq Ids with virus_tax_id = 38018:
- Tax ID: 38018 Modify
- Public page: 38018

Save Check

☐ Curated

[\[Change Log\]](#)

Add

Host Tax ID: (required. Multiple Host Tax ID)

PubMed ID:

Reference:

Comment:

Note:

利用状況

トーゴーの日シンポジウム2021

年間訪問者数（ユニークIPアドレス）

2018年度 113,024人

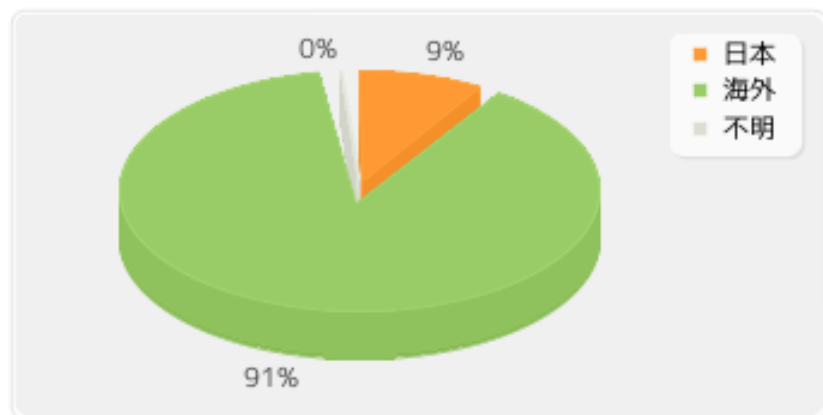
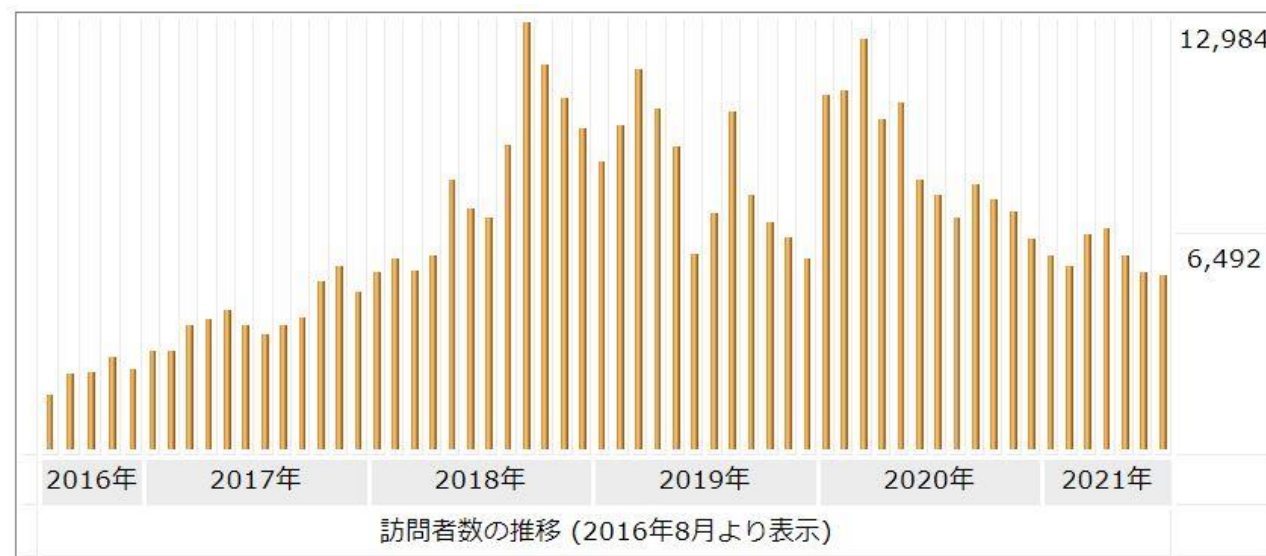
2019年度 104,083人

2020年度 90,977 人

月平均 8,000人以上

合計 408,558人

（2016年7月リリース以来）



国内外のアクセスページ数比率（2020年度）

海外訪問者91%

日本国内 9%

- ウイルス名、生物名の日本語化
- セグメントウイルスの情報の充実化



本研究はJSPS科研費 研究成果公開促進費(20HP7003)、基盤研究B(18H02279)、新学術領域研究(16H06437)、ライフサイエンスデータベース統合推進事業 統合化推進プログラム(21-171023841)の助成を受けたものです。

○朱 梦迪、緒方 博之

京都大学化学研究所バイオインフォマティクスセンター