

# 38 生物世界とウイルス世界を繋ぐデータベース

トーゴーの日シンポジウム2020

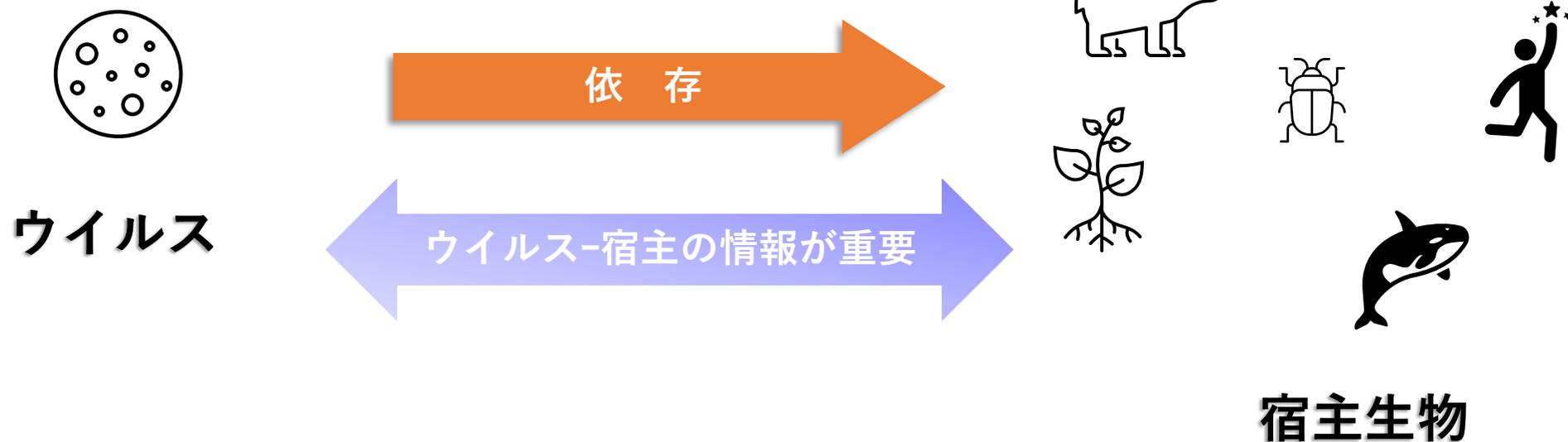
○朱梦迪、緒方博之

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—宿主データベース開発の背景

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

## 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., Barbalescu,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., Barbalescu,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 AY37928.
            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
     source           1..9472
                     /organism="Human endogenous retrovirus K113"
                     /proviral
                     /mol_type="genomic DNA"
                     /host="Homo sapiens"
                     /db_xref="taxon:166122"
     LTR              1..968
     misc_feature     1112..6746
     gene             6451..8550
                     /locus_tag="O179_ori"
```

```
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
- ◆ 信頼性

宿主はフリーテキスト  
大規模処理に不便

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; Bamfordvirae; 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 differ
            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 Biotechnolog
            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, C
COMMENT    PROVISIONAL REFSEQ: This record has not yet been subject
            NCBI review. The reference sequence is identical to KM26
            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
     source           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             Location/Qualifiers
     source           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

```
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
OH 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あるが、  
宿主の登録率が13%**

# ウイルス—宿主データベース

**Virus-Host DB**

About | Statistics | Browse | Virus Index | Host Index | Feedback

Search: Influenza

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]

**Home Page**

**Selected viral families**

- Adenoviridae
- Papillomaviridae
- Herpesviridae
- Baculoviridae
- Poxviridae
- Parvoviridae
- Retroviridae
- Reoviridae
- Picomaviridae
- Flaviviridae
- Caliciviridae
- Coronaviridae
- Bunyaviridae
- Orthomyxoviridae
- Paramyxoviridae

**Annotated viruses** Last Update: August 14, 2020

Viral classification	No. of uniq tax id	Host classification	No. of uniq tax id
dsRNA	445	Bacteria	5
		Eukaryota	437
		Archaea	0
		Viruses	0
		Unassigned/Other	4
dsDNA	3834	Bacteria	2593
		Eukaryota	1140
		Archaea	85
		Viruses	7
		Unassigned/Other	13
ssRNA	3272	Bacteria	18
		Eukaryota	3207
		Archaea	0
		Viruses	4
		Unassigned/Other	48
ssDNA	1738	Bacteria	119
		Eukaryota	1465
		Archaea	1
		Viruses	1
		Unassigned/Other	153
Retrovirus	40	Bacteria	0
		Eukaryota	39
		Archaea	0
		Viruses	0
		Unassigned/Other	1
Satellite virus and Virophage	324	Bacteria	0
		Eukaryota	295
		Archaea	0
		Viruses	44
		Unassigned/Other	20
Viroid	66	Bacteria	0
		Eukaryota	66
		Archaea	0
		Viruses	0
		Unassigned/Other	0
Other	1418	Bacteria	18
		Eukaryota	574
		Archaea	15
		Viruses	0
		Unassigned/Other	1045

**RefSeq** (release 201, July 6, 2020)  
**GenBank** (release 238.0, June 15, 2020)

**Viral entries** (No. of sequence accessions)

15699

## Hop stunt viroid - cucumber

<b>Scientific Name</b>	Hop stunt viroid - cucumber [TAX:12902]
<b>Lineage</b>	Viruses; Pospiviroidae; Hostuviroid; Hop stunt viroid
<b>Genome type</b>	Non-segmented
<b>Genbank</b>	X07405 Cucumber pale fruit viroid (CPFV) RNA.

ウイルス情報

## Known hosts (2)

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

宿主情報

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

# ウイルス—宿主データベース



## 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	Virus lineage	Host name	Host lineage
African swine			
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
African swine fever virus Georgia 2007/1 [TAX:874269]	Viruses; Varidnaviria; Bamfordvirae; Nucleocytoviricota; Pokkesviricetes; Asfuvirales Asfarviridae; Asfivirus; African swine fever virus	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



# アクセス方法・特徴・機能

- **Webでブラウズ（検索）できる**  
(<https://www.genome.jp/virushostdb/>)
- **FTPでファイルをダウンロードできる**  
(<ftp://ftp.genome.jp/pub/db/virushostdb/>)
- **BLAST検索がGenomeNetで可能**  
(<https://www.genome.jp/tools/blast/>)



**GenomeNet のBLAST検索画面へ**

# アクセス方法・特徴・機能

## GenomeNetでのBLAST検索画面

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



### BLAST Search

BLAST FASTA KEGG2

Compute Clear

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

Sequence ID:  (Example) nja\_M3\_1841

Local file name:  参照...

Sequence data: 

```
ctcaactaga gacactcttg agcatctcct cctattaatt ggagacatta
tataggtgtc tctaaatggc attcttgtaa taagttgaac ttttaattga
attaaaaggc tcaaaaggct
```

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



### BLASTN Search Result

Database: genes-nt

DNA sequence database entries related to query - 29 hits

Show alignment

Entry	K number	bits	E-val
Top 10 <input type="button" value="Clear"/> <input type="button" value="Select operation"/> <input type="button" value="Exec"/>			
<input checked="" type="checkbox"/> salt:AO058_05500 peptide chain release factor 1	K15034	42.8	0.40
<input checked="" type="checkbox"/> hmg:101236616 uncharacterized LOC101236616	K09534	41.0	1.4
<input checked="" type="checkbox"/> mesq:C7H62_0080 Outer membrane vitamin B12 receptor BtuB	K16092	39.2	4.9
<input checked="" type="checkbox"/> seon:BWZ22_07880 DNA repair protein RadA	K04485	39.2	4.9
<input checked="" type="checkbox"/> ndo:DDD_2330 polyphosphate kinase		39.2	4.9
<input checked="" type="checkbox"/> camu:CA2015_3802 TonB-dependent receptor plug		39.2	4.9
<input checked="" type="checkbox"/> salq:SYNTR_2085 Glutaminyl-tRNA synthetase	K01886	39.2	4.9
<input checked="" type="checkbox"/> cbj:H04402_01205 glycine betaine transporter OpuD	K05020	39.2	4.9
<input checked="" type="checkbox"/> cbm:CBF_1189 choline/carnitine/betaine transporter	K05020	39.2	4.9
<input checked="" type="checkbox"/> cbf:CLI_1217 choline/carnitine/betaine transporter	K05020	39.2	4.9
<input type="checkbox"/> cbb:CLD_3432 choline/carnitine/betaine transporter	K05020	39.2	4.9
<input type="checkbox"/> cbl:CLK_0571 choline/carnitine/betaine transporter	K05020	39.2	4.9
<input type="checkbox"/> cbh:CLC_1180 choline/carnitine/betaine transporter	K05020	39.2	4.9
<input type="checkbox"/> cba:CLB_1168 choline/carnitine/betaine transporter	K05020	39.2	4.9
<input type="checkbox"/> cbo:CBO1131 opuD; choline/carnitine/betaine transporter	K05020	39.2	4.9
<input type="checkbox"/> ehr:EHR_03820 transcriptional regulator, RofA familiy		39.2	4.9



### ● 他のデータベースに利用 (GLOBI、KEGG)

**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		
	<a href="#">Taxonomy</a>		
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		



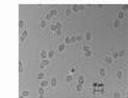
[about](#) [blog](#) [browse](#) [contribute](#) [data](#) [datasets](#) [search](#) [references](#) [日本語](#)

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](#).



has pathogen *Zygosaccharomyces bailii* virus Z

*Zygosaccharomyces bailii*

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/virusostdb/virusostdb.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.



# アップデート

- 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

年間訪問者数（ユニークIPアドレス）

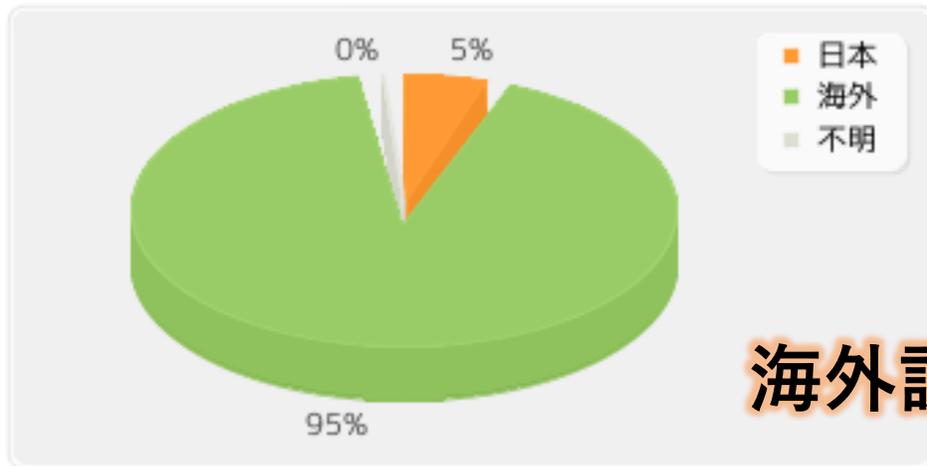
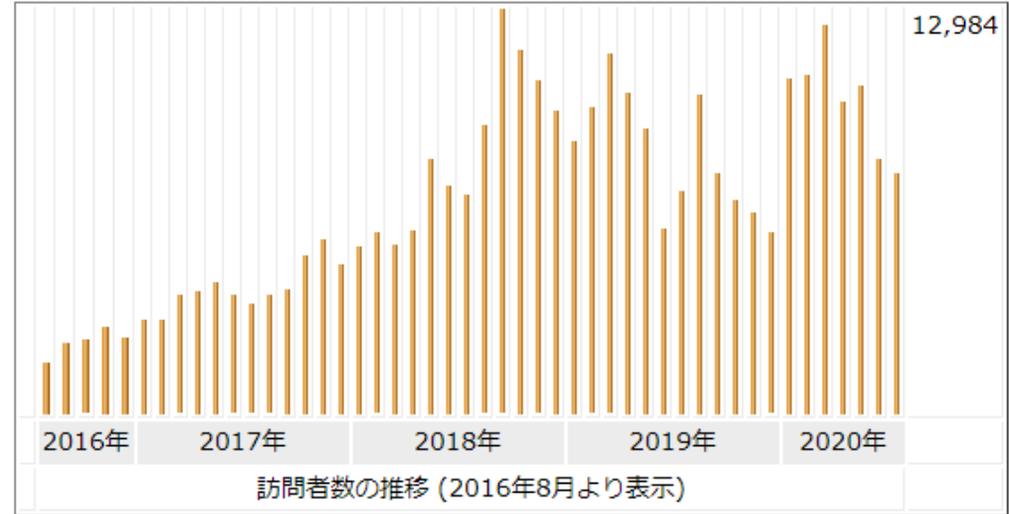
2017年度 55,186 人

2018年度 113,024人

2019年度 104,083人

月平均訪問者数 9,000人以上

2016年7月リリース以来、合計330,920人  
（2020年8月時点）



海外訪問者95%

国内外のアクセスページ数比率（2020年度）

# 今後の課題

- 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
- ウイルスのゲノムのタイプの情報 (dsDNA, ssDNA, dsRNA, etc)



# 謝辞

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)の助成を受けたものです。

