

User's guide



I- Searching and viewing a specific PHROG

PHROGS Prokaryotic virus Remote Homologous Groups

you can take a look at the whole list of PHROGs

Welcome to the Prokaryotic Virus Remote Homologous Groups database (aka. PHROGS)

PHROG : families of prokaryotic virus proteins clustered using remote homology. Terzian P*, Olo Ndela E*, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F.

This database contains 38,880 PHROGs (protein orthologous groups) containing 868,340 proteins from complete genomes of viruses infecting bacteria or archaea (2,318 from RefSeq and 2,669 from GenBank, april 2018), in addition to 12,498 curated prophages derived from cultivated microbial isolates (Roux et al., 2015).

only one standardized annotation was attributed to each PHROG (using RefSeq annotations, and comparison of each PHROG to Pfam, UNIPROT, KEGG and the ACLAME database)

This website provides access to

Phrog-Table

Virus-Table

PHROGs

News

- all prokaryotic virus genomes from the viruses table and select one to see its taxonomy, list of proteins, genomic map, etc...
- all PHROGs from the PHROGs table and select one to see its annotation, list of proteins, multiple alignment, comparison results to Pfam, Uniprot, KEGG, etc...

Viruses and PHROGs can also be access using search tools below.



PHROGS Prokaryotic virus Remote Homologous Groups

Welcome to the Prokaryotic Virus Remote Homologous Groups database (aka. PHROGS)

PHROG : families of prokaryotic virus proteins clustered using remote homology. Terzian P*, Olo Ndela E*, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F.

This database contains 38,880 PHROGs (protein orthologous groups) containing 868,340 proteins from complete genomes of viruses infecting bacteria or archaea (2,318 from RefSeq and 2,669 from GenBank, april 2018), in addition to 12,498 curated prophages derived from cultivated microbial isolates (Roux et al., 2015).

only one standardized annotation was attributed to each PHROG (using RefSeq annotations, and comparison of each PHROG to Pfam, UNIPROT, KEGG and the ACLAME database)

This website provides access to

Virus-Table Phrog-Table

News

- all prokaryotic virus genomes from the viruses table and select one to see its taxonomy, list of proteins, genomic map, etc...
- all PHROGs from the PHROGs table and select one to see its annotation, list of proteins, multiple alignment, comparison results to Pfam, Uniprot, KEGG, etc...

Viruses and PHROGs can also be access using search tools below.

Find a PHROG of interest by searching in :

PHROG annotation terms	neck	Search
Refseq terms	neck	Search
Pfam domain number or annotation	PF05069.12	Search
KEGG Orthology or annotation	K14059	Search
GO term ID or annotation	lysozyme	Search

Or, if you already know the ID of your phrog you can go to its page directly here:

514

PHROG number

Go

Let's have a look at the whole list of PHROGs : The PHROG-Table

Phrog Table

The 38,880 PHROGs are listed below

The whole list (~10Mb) is send to your browser so this can take a few seconds (but then, filtering and searching the table will be fast)

Copy CSV Excel PDF Show 10 v entries

								Search:	
#phrog	hostDomain	Avg_#AA	#CoIMSA	#prot	Annotation	RefSeq	Pfam_hit	GO_hit	KO_hit
phrog_1	Bacteria:5828;Archaea:50	368	2316	5879	integrase	integrase [n=889];	PF00589.21 : Phage_integ	GO:0003677 DNA binding ;	K14059 integrase ; K0476
phrog_2	Bacteria:5770;Archaea:23	506	3767	5795	terminase large subunit	terminase large sub	PF03354.14 : Terminase_1	GO:0016787 hydrolase act	K06909 phage terminase I
phrog_3	Bacteria:3830;unknown:2	75	997	3832	transcriptional repressor	hypothetical protei	PF13744.5 : HTH_37 ; PF	GO:0003677 DNA binding ;	K07726 putative transcri
phrog_4	Bacteria:3265	230	1098	3265	transcriptional repressor	hypothetical protei	PF08667.9 : BetR ; PF14	GO:0051259 protein compl	K01356 repressor LexA [E
phrog_5	Bacteria:3214;Archaea:10	135	599	3224	neck protein Ne1	hypothetical protei	PF05521.10 : Phage_H_T_j	NA	NA
phrog_6	Bacteria:3188;Archaea:14	351	1677	3202	baseplate protein	hypothetical protei	PF04865.13 : Baseplate_J	NA	NA
phrog_7	Bacteria:3145	170	1541	3145	endolysin	lysozyme [n=162]; e	PF00959.18 : Phage_lysoz	GO:0003796 lysozyme acti	K01185 lysozyme [EC:3.2
phrog_8	Bacteria:3060;Archaea:7	131	1182	3067	transcriptional regulator	hypothetical protei	PF10078.8 : DUF2316 ; P	GO:0006355 regulation of	K07729 putative transcri

For each PHROG the table will show the following headers :



Search	Search	Search	Search	Search	Search	Search	Search	Search	Search
phrog_10	Bacteria:2870	375	1712	2870	major head protein	major capsid protei	PF04586.16 : Peptidase_S	NA	K06904 uncharacterized p
phrog_9	Bacteria:2977	557	2362	2977	terminase large subunit	terminase large sub	PF03354.14 : Terminase_1	GO:0016787 hydrolase act	NA
phrog_8	Bacteria:3060;Archaea:7	131	1182	3067	transcriptional regulator	hypothetical protei	PF10078.8 : DUF2316 ; P	GO:0006355 regulation of	K07729 putative transcri
phrog_7	Bacteria:3145	170	1541	3145	endolysin	lysozyme [n=162]; e	PF00959.18 : Phage_lysoz	GO:0003796 lysozyme acti	K01185 lysozyme [EC:3.2
phrog_6	Bacteria:3188;Archaea:14	351	1677	3202	baseplate protein	hypothetical protei	PF04865.13 : Baseplate_J	NA	NA
phrog_5	Bacteria:3214;Archaea:10	135	599	3224	neck protein Ne1	hypothetical protei	PF05521.10 : Phage_H_T_j	NA	NA

Showing 1 to 10 of 38,880 entries

Previous 1 2 3 4 5 ... 3888 Next

Each column of this table is searchable

PHROGs News Virus-Table Phrog-Table Suggestions Documentation

PHROGS Prokaryotic virus Remote Homologous Groups

Welcome to the Prokaryotic Virus Remote Homologous Groups database (aka. PHROGS)

PHROG : families of prokaryotic virus proteins clustered using remote homology. Terzian P*, Olo Ndela E*, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F.

This database contains 38,880 PHROGs (protein orthologous groups) containing 868,340 proteins from complete genomes of viruses infecting bacteria or archaea (2,318 from RefSeq and 2,669 from GenBank, april 2018), in addition to 12,498 curated prophages derived from cultivated microbial isolates (Roux et al., 2015).

only one standardized annotation was attributed to each PHROG (using RefSeq annotations, and comparison of each PHROG to Pfam, UNIPROT, KEGG and the ACLAME database)

This website provides access to

- all prokaryotic virus genomes from the viruses table and select one to see its taxonomy, list of proteins, genomic map, etc...
- all PHROGs from the PHROGs table and select one to see its annotation, list of proteins, multiple alignment, comparison results to Pfam, Uniprot, KEGG, etc...

Go

Viruses and PHROGs can also be access using search tools below.

Find a PHROG of interest by searching in :

PHROG annotation terms	neck	Search
Refseq terms	neck	Search
Pfam domain number or annotation	PF05069.12	Search
KEGG Orthology or annotation	K14059	Search
GO term ID or annotation	lysozyme	Search

514

Or, if you already know the ID of your phrog you can go to its page directly here:

PHROG number

You can also find a PHROG with this menu

Find a PHROG of interest by searching in :

PHROG annotation terms	neck	Search	
Refseq terms	neck	Search	
Pfam domain number or annotation	PF05069.12	Search	
KEGG Orthology or annotation	K14059	Search	
GO term ID or annotation	lysozyme	Search	
		Write some text and search in the different terms or annotation	

different terms or annotation for example "lysozyme" in its Gene Ontology annotation All PHROGs containing here "lysozyme" in its GO annotation are presented in a table similar to the whole list of PHROGs

List of phrogs annotated with the GO term : "lysozyme"

Copy CSV Excel PDF

Search:

#phrog ^{↑↓}	Domain 斗	#MS ^{↑↓}	#colMSA 🛝	#prot ↑↓	#annot ^{†⊥}	Phrog_annot	RefSeq 1	pfam î↓	KEGG 11	GO îl
phrog_7	Bacteria:3145	944	1541	3145	557	endolysin	lysozyme [n=162]; endolysin [n=146	PF00959.18&Phage_lysozyme ; Ph	K01185&E3.2.1.17&lysozyme	GO:0003796 lysozyme activity !!!GO
phrog_291	Bacteria:424	545	1118	424	134	endolysin	endolysin [n=50]; lysin [n=46]; ly	PF08924.10&DUF1906 ; Domain of	K07273&acm&lysozyme&&a	GO:0016998 cell wall macromolecule
phrog_2083	Bacteria:72	940	1176	72	41	internal virion lysozyme motif	hypothetical protein [n=29]; inter	PF00959.18&Phage_lysozyme ; Ph	K01185&E3.2.1.17&lysozyme	GO:0009253 peptidoglycan catabolic
phrog_2473	Bacteria:60	161	162	60	0	NA	hypothetical protein [n=2];	PF00959.18&Phage_lysozyme ; Ph	NA	GO:0009253 peptidoglycan catabolic
phrog_3084	Bacteria:47	559	692	47	39	endolysin	lysin A [n=23]; lysin [n=7]; lysin…	PF01183.19&Glyco_hydro_25; Gl	NA	GO:0009253 peptidoglycan catabolic
phrog_3393	Bacteria:42	376	496	42	28	endolysin	lysin [n=13]; endolysin [n=13]; hy	PF01183.19&Glyco_hydro_25; Gl	K01447&E3.5.1.28A, cwlA, xlyA,	GO:0016998 cell wall macromolecule
phrog_3572	Bacteria:40	653	653	40	40	endolysin	peptidase M23 [n=40];	PF16754.4&Pesticin ; Bacterial	NA	GO:0009401 phosphoenolpyruvate- dep
phrog_3617	Bacteria:40	486	486	40	0	NA	hypothetical protein [n=40];	PF00959.18&Phage_lysozyme ; Ph	NA	GO:0005524 ATP binding !!!GO:00469
phrog_4290	Bacteria:32	973	1058	32	12	lysozyme domain- containing protein	hypothetical protein [n=20]; lysoz	PF00959.18&Phage_lysozyme ; Ph	NA	GO:0009253 peptidoglycan catabolic
phrog_5963	Bacteria:21	156	163	21	2	holin	hypothetical protein [n=19]; holin	PF01183.19&Glyco_hydro_25; Gl	NA	GO:0016998 cell wall macromolecule

When clicking on any PHROG ID, a summary page will appear, for instance, if PHROG_1 is selected...

#phrog	hostDomain	Avg_#AA	#ColMSA	#prot	Annotation	RefSeq	Pfam_hit	GO_hit	KO_hit
phrog_1	Bacteria:5828;Archaea:50	368	2316	5879	integrase	integrase [n=889];	PF00589.21 : Phage_integ	GO:0003677 DNA binding ;	K14059 integrase ; K0476
phrog_2	Bacteria:5770;Archaea:23	506	3767	5795	terminase large subunit	terminase large sub	PF03354.14 : Terminase_1	GO:0016787 hydrolase act	K06909 phage terminase I
phrog_3	Bacteria:3830;unknown:2	75	997	3832	transcriptional repressor	hypothetical protei	PF13744.5 : HTH_37 ; PF	GO:0003677 DNA binding ;	K07726 putative transcri
phrog_4	Bacteria:3265	230	1098	3265	transcriptional repressor	hypothetical protei	PF08667.9 : BetR ; PF14	GO:0051259 protein compl	K01356 repressor LexA [E
phrog_5	Bacteria:3214;Archaea:10	135	599	3224	neck protein Ne1	hypothetical protei	PF05521.10 : Phage_H_T_j	NA	NA
phrog_6	Bacteria:3188;Archaea:14	351	1677	3202	baseplate protein	hypothetical protei	PF04865.13 : Baseplate_J	NA	NA
phrog_7	Bacteria:3145	170	1541	3145	endolysin	lysozyme [n=162]; e	PF00959.18 : Phage_lysoz	GO:0003796 lysozyme acti	K01185 lysozyme [EC:3.2

PHROGs News Virus-Table Phrog-Table Suggestions Documentation

PHROGS Prokaryotic virus Remote Homologous Groups

Welcome to the Prokaryotic Virus Remote Homologous Groups database (aka. PHROGS)

PHROG : families of prokaryotic virus proteins clustered using remote homology. Terzian P*, Olo Ndela E*, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F.

This database contains 38,880 PHROGs (protein orthologous groups) containing 868,340 proteins from complete genomes of viruses infecting bacteria or archaea (2,318 from RefSeq and 2,669 from GenBank, april 2018), in addition to 12,498 curated prophages derived from cultivated microbial isolates (Roux et al., 2015).

only one standardized annotation was attributed to each PHROG (using RefSeq annotations, and comparison of each PHROG to Pfam, UNIPROT, KEGG and the ACLAME database)

This website provides access to

- all prokaryotic virus genomes from the viruses table and select one to see its taxonomy, list of proteins, genomic map, etc...
- all PHROGs from the PHROGs table and select one to see its annotation, list of proteins, multiple alignment, comparison results to Pfam, Uniprot, KEGG, etc...

Viruses and PHROGs can also be access using search tools below.

Find a PHROG of interest by searching in :

PHROG annotation terms	neck	Search
Refseq terms	neck	Search
Pfam domain number or annotation	PF05069.12	Search
KEGG Orthology or annotation	K14059	Search
GO term ID or annotation	lysozyme	Search

Or, if you already know the ID of your phrog you can go to its page directly here:

514

PHROG number

Go

You will also get the PHROG summary page when browsing by PHROG number

On each PHROG page, you will have the list of all proteins of the PHROG, with their NCBI ID, their NCBI annotation, name of their virus and its taxonomy

phrog_1 is made of 5879 protein sequences and is annotated as "integrase"

Average length of proteins	Number of columns in the MSA	Number of conserved columns in the MSA
368.27	2316	332

List of all proteins of this PHROG

Search:

PHROG prot	NCBI prot ID	NCBI prot annotation	Virus ID (click to view)	Virus name	Virus taxonomy	†1
p428256	no NCBI prot ID	no annotation	VI_04636	Bacilli_gi_484970147	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; unclassified Siphoviridae.	
p267170	no NCBI prot ID	no annotation	VI_05328	Bacilli_gi_544790345	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; unclassified Siphoviridae.	
p154024	no NCBI prot ID	no annotation	VI_04550	Bacilli_gi_485066273_238_40694	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; unclassified Siphoviridae.	
KX833905_p1	APC44627.1	hypothetical protein	KX833905	Lactococcus phage PLg-TB25	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; unclassified Siphoviridae.	
p166464	no NCBI prot ID	no annotation	VI_00836	Bacilli_gi_385836969_622724_662135	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; unclassified Siphoviridae.	
110100	1001 110	·	1/1 00700	B 18 - 51070000 000051 071010	SP	

You will also find 4 tables with information on colocalizing PHROGs, families from PHROGs, PFAM that are similar to the current PHROG, with e-value, alignment position, pathways, the best scoring hit and the number of hits when possible The annotations found in the NCBI proteins of this PHROG are listed in the fourth table

PHROGs colocalized to this PHROG

			565	Gearch.				
PHROG 1	#Prot	Annotation	#Neighbors	Probability 1	Score 1			
phrog_147	653	transcriptional regulator	548	0	1000			
phrog_539	254	excisionase	236	0	1000			
phrog_629	224	no annotation	213	0	1000			
phrog_2240	67	no annotation	67	0	1000			
phrog_2842	51	no annotation	51	0	1000			
phrog_7259	16	no annotation	30	0	1000			
phrog_8946	12	no annotation	20	0	1000			
phrog_7835	15	no annotation	19	0	1000			

PHROGs similar to this PHROG

			Search:	
PhrogHits 1	Phrog's Definition	Probability 14	Evalue 11	posQ/posS
phrog_216	integrase	99.8	6.2E-26	149-322 / 6-178
phrog_7244	integrase	99.4	1.3E-16	45-226 / 12-188
phrog_7685	integrase	99.0	1.4E-13	237-326 / 10-102
phrog_17497	no annotation	98.3	7.8E-10	236-328 / 10-111
phrog_22184	no annotation	98.3	4.4E-10	41-202 / 64-243
phrog_35951	no annotation	98.3	4.2E-10	268-326 / 19-76
phrog_23271	no annotation	98.0	5.9E-09	268-313 / 7-51
phrog_7269	integrase	97.8	2.6E-08	268-323 / 32-86

PFAMs similar to this PHROG

		Search:				
ID 11	Definition 1	Evalue	posQ/posS			
PF16795.4	Archaeal phage integrase	1.4E-06	303-467 / 3-157			
PF13356.5	Arm DNA-binding domain	0.0009	138-179 / <mark>24-</mark> 67			
PF16787.4	Centromere DNA-binding protein complex CBF3 subunit, domain 2	2.8E-06	308-470 / 16-223			
PF12835.6	Integrase	1.5E-05	298-436 / 8-149			
PF00589.21	Phage integrase family	2.5E-09	303-468 / 2-170			
PF13102.5	Phage integrase SAM-like domain	0.00027	194-282 / 1-95			
PF12834.6	Phage integrase, N-terminal	0.0035	211-279 / 16-88			
PF13495.5	Phage integrase N-terminal SAM-like domain	0 0021	211-280 / 13-83			

Annotations of the NCBI proteins of this PHROG

889:integrase	41:hypothetical protein	30:protein of an unannotat	29:phage integrase
19:site-specific recombinase	11:integrase/site specific r	9:tyrosine integrase	8:integrase (Y-int)
7:integrase protein	7:mobile element protein	6:integrase family protein	6:Int
5:integrase domain protein	5:integrase/recombinase	5:Integrase	5:site-specific integrase
4:site-specific tyrosine rec	4:phage integrase family)	4:integrase Int	4:Y-integrase
4:integrase (Y-Int)	4:recombinase/integrase	4:site specific recombinas)	3:prophage integrase
3:shufflon-specific DNA re	3:integrase-recombinase	3:integrase Y-int	3:phage integrase family s
3:tyrosine recombinase	3:Phage integrase	2:site specific recombinase	2:integrase family site spe
2:cyclization recombinase	2:site-specific tyrosine rec	(2:XerC/XerD family integr)	2:intergrase
2:XerD-like integrase	2:site-specific recombinas)	(2:DNA integration/recomb)	2:XerC/D family recombin
2:integrase, (Y-int)	2:integrase (y-int)	2:bacteriophage integrase	2:Tyrosine integrase
2:Integrase (Y-int)	2:integrase-like protein	2:recombinase XerD	1:Tyr integrase
1:probable phage-family i	1:puative integrase protein	1:tyrosine recombinase X	1:integrase (Tyr)
1:phage integrase/site-sp	1:DNA integration/recomb	1:helix-turn-helix domain	1:DNA integration/recomb

The last two tables show KEGG orthologous groups and Gene Ontology terms from Uniprot proteins that are similar to the current PHROG, also with e-value, alignment position, pathways, the best scoring hit and the number of hits when possible

PFAMs similar to this PHROG

ID 11	Definition 1	Evalue	posQ/posS
PF16795.4	Archaeal phage integrase	1.4E-06	303-467 / 3-157
PF13356.5	Arm DNA-binding domain	0.0009	138-179 / 24-67
PF16787.4	Centromere DNA-binding protein complex CBF3 subunit, domain 2	2.8E-06	308-470 / 16-223
PF12835.6	Integrase	1.5E-05	298-436 / 8-149
PF00589.21	Phage integrase family	2.5E-09	303-468 / 2-170
PF13102.5	Phage integrase SAM-like domain	0.00027	194-282 / 1-95
PF12834.6	Phage integrase, N-terminal	0.0035	211-279 / 16-88
PF13495.5	Phage integrase N-terminal SAM-like domain	0 0021	211-280 / 13-83

Search

Search

KEGG orthologous groups similar to this PHROG

ID 11	Name : Definition	Pathways ↑↓	BestScore	#Hits
K14059	int : integrase	no pathways	805	365
K04763	xerD : integrase/recombinase XerD	no pathways	623	45
K03733	xerC : integrase/recombinase XerC	no pathways	447	22

Annotations of the NCBI proteins of this PHROG

889:integrase	41:hypothetical protein	30:protein of an unannotat	29:phage integrase
19:site-specific recombinase	11:integrase/site specific r	9:tyrosine integrase	8:integrase (Y-int)
7:integrase protein	7:mobile element protein	6:integrase family protein	6:Int
5:integrase domain protein	5:integrase/recombinase	5:Integrase	5:site-specific integrase
4:site-specific tyrosine rec)	4:phage integrase family	4:integrase Int	4:Y-integrase
4:integrase (Y-Int)	4:recombinase/integrase	4:site specific recombinas	3:prophage integrase
3:shufflon-specific DNA re	3:integrase-recombinase	3:integrase Y-int	3:phage integrase family s
3:tyrosine recombinase	3:Phage integrase	2:site specific recombinase	2:integrase family site spe
2:cyclization recombinase	2:site-specific tyrosine rec)	2:XerC/XerD family integr	2:intergrase
2:XerD-like integrase	2:site-specific recombinas)	(2:DNA integration/recomb)	2:XerC/D family recombin
2:integrase, (Y-int)	2:integrase (y-int)	2:bacteriophage integrase	2:Tyrosine integrase
2:Integrase (Y-int)	2:integrase-like protein	2:recombinase XerD	1:Tyr integrase
1:probable phage-family i)	1:puative integrase protein	1:tyrosine recombinase X	1:integrase (Tyr)
1:phage integrase/site-sp)	1:DNA integration/recomb	1:helix-turn-helix domain	1:DNA integration/recomb

GO terms in Uniprot proteins similar to this PHROG

	Search:
ID	11 Definition 11
GO:0003677	DNA binding
GO:0006310	DNA recombination
GO:0015074	DNA integration

At the bottom of the page you will find the multiple sequence alignment of all proteins of the current PHROG, embedded in an interactive way, to allow you to scroll and move through it

	orthologous grou	ups simi	lar to thi	s PHR	OG Search:	(GO terms in l	Jniprot protein	s similar to this PH	ROG Search:	
	Name : Definition	Pathways 1	BestScore	#Hits 斗			ID	1	Definition		
K14059	int : integrase	no pathways	805	365			GO:0003677		DNA binding		
K04763	xerD : integrase/recombinase	no pathways	623	45			GO:0006310		DNA recombination		
100700	XerD	an anthursus	447	22			GO:0015074		DNA integration		
NU3733	XerC	no patriways		22							
Import (ID Label 1 p4282 2 p2671	Sorting Filter Selection Vis.elemen 452. 454. 456. 458. 460. 462 56 - <	s Color scheme , 464, 466, 468, 4 Q E T E	Extras Export 70. 472. 474. 476	Help . 478. 480. 482.	484, 486, 488, 490, 492, 494, 496, 498 K H -	1. 500. 502. 504. 50 V K - L S	6. 508. 510. 512. 514. 516. :	518, 520, 522, 524, 526, 528, 53	0. 532. 534. 536. 538. 540. 542. 544. 546 , A D V S N I T , N E G S D I T	. 548. 550. 552. 554. 556. 558. 560. 562	2.564.56
3 p1540 4 KX833 5 p1664	24) E		K I I) К Н	V K				F V D L - Y D	
5 prooi		T E		K I I	К Н	V K L S			A D V S N I T	F V D L - Y D F W E L - Y E F W E L - Y E F V D L - Y D	
6 p1424 7 p1443	69	TE		K I I K I M K I A) K H	- V - K - L - S - L - N	K G		A D V S N I T	F V D L Y B C B C C C C C C C C C C C C C C C C	
6 p1424 7 p1443 8 p1961 9 p1945 10 p1940	69	TE NN XK XK XK		K I I K I I K I I I L I I L I	I K I	V K - L - S - L - N - E E E K R L E E E K R L E F F K R L	K G	E - GV - I - P E - GV - I - P E - GV - I - P	A D V S N I T A D V S N I T N E G S D I T D V K N O N L T D V K N O N L T N D K T N D K T N D K T	F V D L Y - D F W E L Y - E F W E L Y - E F V D L Y - D F K E L Y - D L E D M I - K L E D M I - K	
6 p1424 7 p1443 8 p1961 9 p1945 10 p1940 11 p1958 12 p1957 13 p1958	69	T E		K I I K I K I I L I L I L I L		V K - L N - L N - E E E K R L - E E E K R L	K G -		A D V S N I T A D V S N I T N E G S D I T D V K N O N L T D V K N O N L T N D K T		
6 p1424 7 p1443 8 p1961 9 p1945 10 p1940 11 p1958 12 p1957 13 p1958 14 NC 01 15 p1951 16 p1949	69 46 77 18 19 19 10 10 10 10 10 11 10 11 11 11	T E		K I I K I I K I I I L I L I L I L I L I L I L I L I L I L	N K H D A D A D Y D Y D Y D Y D Y D Y D Y D Y	- V - K - L - N - E E E K R L - E E E K R L			A D V S N I T T A D V S N I T T N E G S D I T D V K N Q N L T D V K N Q N L T N D K T	F V D L Y D F V L Y E F W L Y E F W L Y E F W L Y D F K L Y D F K L Y D F K L Y D L E D M I K L E D M I K L E D M I K L E D M I K L E D M I K L E M I K L	
6 p1424 7 p1443 8 p1961 9 p1945 10 p1940 11 p1958 12 p1957 13 p1958 14 NC 01 15 p1951 16 p1949 17 p1941 18 p1953 19 p1944	69 I K I 46 I K I 23 I K I 10 I K I 11 I K I 12 I K I 13 I K I	T E		K I K K I K K I K I L I L I L I L I L I L I L I L I L I L		V K V K V K V N	K G K G K G K G K G K G K N C N L K N C N L K N C N L K N C K N C C K N C K N C C C C C C C C C C C C C	E G V - I - P G V	A D V S N I T A D V S N I T N E G S D I T D V K N O N L T D V K N O N L T N D K T		
6 p1424 7 p1443 8 p1961 9 p1945 10 p1940 11 p1958 12 p1957 13 p1958 14 NC 01 15 p1951 16 p1949 17 p1941 18 p1953 19 p1944 20 p8677 21 p1948 22 NC 00	69 46 77 16 77 17 17 17 17 17 17 17 17 17	T E		<pre>K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K I I K K I I K K I I K K I I K K K K</pre>	0 K H 0 A	- V . K - L . N - E E E K R L - E E K R L - E E E K R L - E E K R L - E E K R L - E E K R L K - E K R L - E E K R L K - E K - E K R L K - E K - E K R L K - E	K G K K	E G V - I - P E - G G V - I - P E - G V - I - P	A D V S N I T A D V S N I T N E G S D I T D V K N Q N L T N D V K N Q N L T N D K T N	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	
3 p1024 7 p1443 8 p1961 10 p1940 11 p1958 12 p1957 13 p1958 14 NC 01 15 p1951 16 p1949 17 p1944 19 p1948 22 NC 00 23 p1969 24 p1963 25 p1951 16 p1949 17 p1944 22 NC 00 23 p1969 24 p1963 25 p1953 25 p1954 25 p1954 25 p1954 25 p1954 25 p1954 25 p1954 25 p1954 26 p1954 27 p1955 27 p1955	69	T E		<pre>K I I V</pre>	N H D A D A D A D Y				A D V S N I T T A D V S N I T T N E G S D I T D V K N Q N L T D V K N Q N L T N D K T N	F V D L Y D F V D L Y D F V D L Y D F V D L Y D F V D L Y D F K E L Y D F K E L Y D F K E L Y D L E D M T K L E D M T K L E D M T K L E D M T K L E D M T K L E D M T K L E D M T K L E	

For smaller PHROGs, an overview image of the multiple alignment is available, and allows you to zoom on a specific region on the alignment, here the region corresponds to the gray square (selected by a simple click, hold, glide and release)

12 NC 02299 30	N D C I R R D C M R I R D C M R I N D C I R N D C I R	A - - R A -	I H V I I H V I I P V I I P V I I P V I I P V I I P V I I P V I P V I P V I P V I P V I P V I P V I I P V I P V I I P V I P V I I P V I P V I I P V I I V I	Y K G D N Y G D N Y G D N V R -		T r			
								7981	Zoom area



II- Searching and viewing a specific virus

PHROGs News Virus-Table Charge-Table Suggestions Documentation

PHROGS Prokaryotic virus Remote Homologo

To view all information about a Virus, all you have to do is select one

Welcome to the Prokaryotic Virus Remote Homologous Groups database (aka. PHROGS)

PHROG : families of prokaryotic virus proteins clustered using remote homology. Terzian P*, Olo Ndela E*, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F.

This database contains 38.880 PHROGs (protein orthologous groups) containing 868,340 proteins from complete genomes of viruses infecting bacteria or archaea (2,318 from RefSeq and 2,669 from GenBank, april 2018), in addition to 12,498 curated prophages derived from cultivated microbial isolates (Roux et al., 2015).

only one standardized annotation was attributed to each PHROG (using RefSeq annotations, and comparison of each PHROG to Pfam, UNIPROT, KEGG and the ACLAME database)

This website provides access to :

- all prokaryotic virus genomes from the viruses table and select one to see its taxonomy, list of proteins, genomic map, etc...
- all PHROGs from the PHROGs table and select one to see its annotation, list of proteins, multiple alignment, comparison results to Pfam, Uniprot, KEGG, etc...

Viruses and PHROGs can also be access using search tools below.

Find a PHROG of interest by searching in :

PHROG annotation terms	neck	Search
Refseq terms	neck	Search
Pfam domain number or annotation	PF05069.12	Search
KEGG Orthology or annotation	K14059	Search
GO term ID or annotation	lysozyme	Search
PHROG number	514	Go
Find a virus of interest by searcl	ning in :	
Find a virus of interest by searcl Virus ld	ning in : NC_011048	Search
Find a virus of interest by search Virus ld If you don't know the accession of your vir	NC_011048	Search
ind a virus of interest by search Virus Id If you don't know the accession of your vir Virus name	INC_011048	Search

you can take a look at the whole list of Viruses

or select one by searching with its ID or its name

PHROGS Prokaryotic virus Remote Homologous Groups

Let's have a look at the whole list of viruses : The Virus-Table

Welcome to the Prokaryotic Virus Remote Homologous Groups database (aka. PHROGS)

PHROG : families of prokaryotic virus proteins clustered using remote homology. Terzian P*, Olo Ndela E*, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F.

This database contains 38,880 PHROGs (protein orthologous groups) containing 868,340 proteins from complete genomes of viruses infecting bacteria or archaea (2,318 from RefSeq and 2,669 from GenBank, april 2018), in addition to 12,498 curated prophages derived from cultivated microbial isolates (Roux et al., 2015).

only one standardized annotation was attributed to each PHROG (using RefSeq annotations, and comparison of each PHROG to Pfam, UNIPROT, KEGG and the ACLAME database)

This website provides access to

- all prokaryotic virus genomes from the viruses table and select one to see its taxonomy, list of proteins, genomic map, etc...
- all PHROGs from the PHROGs table and select one to see its annotation, list of proteins, multiple alignment, comparison results to Pfam, Uniprot, KEGG, etc...

Viruses and PHROGs can also be access using search tools below.

og-Table Suggestions Documentation

Virus-Table

Find a PHROG of interest by searching in :

PHROG annotation terms	neck	Search
Refseq terms	neck	Search
Pfam domain number or annotation	PF05069.12	Search
KEGG Orthology or annotation	K14059	Search
GO term ID or annotation	lysozyme	Search

Or, if you already know the ID of your phrog you can go to its page directly here:

514

PHROG number

Go

Virus Table

~ 5,000 complete viral genomes from the NCBI and ~ 12,500 curated prophages previously published (termed "VI_...)

Copy CSV Excel PDF Show 10 ~ entries

												Search:		
Locus 📋	Definition	MolType	MolStructure	Taxonomy	DB_origin	Length	Host name	Host Domain	is.prophage?	nb- Proteins	%GC	#Phrogs	nb- Sing	nb- Paralogue
A3_VLP_circ	Methanococcus voltae A3 chromosome, complete genome.	DNA	linear	Viruses; unclassified archaeal viruses.	NUCCORE	25910	Methanococcus voltae	Archaea	undef	34	27.48	6	28	0
AB045978	Staphylococcus phage phiSLT DNA, complete genome.	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudoviricetes; Siphoviridae; Triavirus.	NUCCORE	42942	Staphylococcus	Bacteria	undef	60	33.31	59	0	1
AB334721	Enterobacteria phage 11 DNA, complete genome, isolate: NBRC 20015.	DNA	circular	Viruses; Monodnaviria; Loebvirae; Hofneiviricota; Faserviricetes; Tubulavirales; Inoviridae; Inovirus; Escherichia virus M13.	PVOG	6407	Enterobacteriaceae	Bacteria	undef	10	40.84	10	0	0
AB426868	Enterobacteria phage P22	DNA	circular	Viruses; Duplodnaviria;	NUCCORE	41660	Salmonella enterica	Bacteria	undef	71	47.01	69	1	1



When clicking on any Virus ID, a summary page will appear

Locus 📋	Definition	MolType	MolStructure	Taxonomy	DB_origin	Length	Host name	Host Domain	is.prophage?	nb- Proteins	%GC	#Phrogs	nb- Sing	nb- Paralogue
A3_VLP_circ	Methanococcus voltae A3 chromosome, complete genome.	DNA	linear	Viruses; unclassified archaeal viruses.	NUCCORE	25910	Methanococcus voltae	Archaea	undef	34	27.48	6	28	0
AB045978	Staphylococcus phage phiSLT DNA, complete genome.	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; Triavirus.	NUCCORE	42942	Staphylococcus	Bacteria	undef	60	33.31	59	0	1
AB334721	Enterobacteria phage f1 DNA, complete genome, isolate: NBRC 20015.	DNA	circular	Viruses; Monodnaviria; Loebvirae; Hofneiviricota; Faserviricetes; Tubulavirales; Inoviridae; Inovirus; Escherichia virus M13.	PVOG	6407	Enterobacteriaceae	Bacteria	undef	10	40.84	10	0	0
AB426868	Enterobacteria phage P22	DNA	circular	Viruses; Duplodnaviria;	NUCCORE	41660	Salmonella enterica	Bacteria	undef	71	47.01	69	1	1

PHROGS Prokaryotic virus Remote Homologous Groups

Welcome to the Prokaryotic Virus Remote Homologous Groups database (aka. PHROGS)

PHROG : families of prokaryotic virus proteins clustered using remote homology. Terzian P*, Olo Ndela E*, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F,

This database contains 38.880 PHROGs (protein orthologous groups) containing 868,340 proteins from complete genomes of viruses infecting bacteria or archaea (2,318 from RefSeq and 2,669 from GenBank, april 2018), in addition to 12,498 curated prophages derived from cultivated microbial isolates (Roux et al., 2015).

only one standardized annotation was attributed to each PHROG (using RefSeq annotations, and comparison of each PHROG to Pfam, UNIPROT, KEGG and the ACLAME database)

This website provides access to :

- all prokaryotic virus genomes from the viruses table and select one to see its taxonomy, list of proteins, genomic map, etc...
- all PHROGs from the PHROGs table and select one to see its annotation, list of proteins, multiple alignment, comparison results to Pfam, Uniprot, KEGG, etc...

Viruses and PHROGs can also be access using search tools below.

Find a PHROG of interest by searching in :

PHROG annotation terms	neck	Search
Refseq terms	neck	Search
Pfam domain number or annotation	PF05069.12	Search
KEGG Orthology or annotation	K14059	Search
GO term ID or annotation	lysozyme	Search
Or, if you already know the ID of you	ur phrog you can go to its page	e directly here:
Or, if you already know the ID of you PHROG number	ur phrog you can go to its page	e directly here:
Or, if you already know the ID of you PHROG number Find a virus of interest by sea Virus Id	sr phrog you can go to its page 514 rching in :	e directly here: Go
Or, if you already know the ID of you PHROG number Find a virus of interest by sea Virus Id	strphrog you can go to its page 514 rching in : NC_011048	e directly here: Go Search
Or, if you already know the ID of you PHROG number Find a virus of interest by sea Virus Id If you don't know the accession of your	rching in : NC_011048 rvirus, use the search box below	e directly here: Go Search

You can also find a Virus with the search boxes here

Find a virus of interest by searching in :

Virus Id

NC_011048 Search

If you don't know the accession of your virus, use the search box below :

Virus name



With this search box, you will have a table listing all viruses with the given key word in their name

For example, when searching "Lambda", you will get the following table, with the same headers as the Virus-table

Copy CSV Excel PDF

Lambda-like

viruses

like viruses

Viruses with Lambda in their name

Search'

Locus 14	Name 1	Definition 1	MolType	MolStructure	Taxonomy	DB_origin ↑	Length 斗	Host name ↑↓	Host Domain	is.prophage?	nb- Proteins	%GC 斗	#Phrogs	nb- Sing ^{↑↓}	nb- Paralogue	
NC_001416	Escherichia virus Lambda	Enterobacteria phage lambda, complete genome.	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; Lambdavirus.	REFSEQ	48502	Escherichia coli	Bacteria	undef	73	49.85	72	0	1	
VI_05505	Uncultured unclassified Lambda-like viruses	Uncultured unclassified Lambda- like viruses	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; Lambdavirus; unclassified Lambdavirus.	VirSorter	23020		Bacteria	no	25	61.09	20	5	0	
VI_05550	Uncultured unclassified Lambda-like viruses	Uncultured unclassified Lambda- like viruses	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; Lambdavirus; unclassified Lambdavirus.	VirSorter	47405		Bacteria	yes	59	60.58	56	3	0	
VI_03013	Uncultured unclassified Lambda-like viruses	Uncultured unclassified Lambda- like viruses	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; Lambdavirus; unclassified Lambdavirus.	VirSorter	45817		Bacteria	yes	55	48.92	54	0	1	
VI_06553	Uncultured unclassified Lambda-like viruses	Uncultured unclassified Lambda- like viruses	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; Lambdavirus; unclassified Lambdavirus.	VirSorter	36162		Bacteria	yes	45	52.28	45	0	0	
VI_06557	Uncultured unclassified Lambda-like viruses	Uncultured unclassified Lambda- like viruses	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Siphoviridae; Lambdavirus; unclassified Lambdavirus.	VirSorter	37154		Bacteria	no	40	50.51	38	0	2	
VI_06529	Uncultured unclassified	Uncultured unclassified Lambda-	DNA	linear	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes;	VirSorter	34849		Bacteria	yes	43	52.25	43	0	0	

Caudovirales; Siphoviridae;

Lambdavirus; unclassified Lambdavirus.

A summary phage will also appear when browsing for a Virus ID on the home page, for example if we search "NC_011048"...

Find a virus of interest by searching in :

Virus Id

NC_011048

If you don't know the accession of your virus, use the search box below :

Virus name

Lambda	Search
--------	--------

Genomic map of Bacillus virus phi29 (NC_011048)



The summary virus page is presented as follows, with...

A genomic map of the virus

More about this genome

0 2k 4k 6k 8k 10k 12k 14k 16k 18k

	5
Locus	NC_011048
Definition	Bacillus phage phi29, complete genome.
Molecule type	DNA
Genome structure	linear
Taxonomy	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Podoviridae; Picovirinae; Salasvirus.
DB of origin	REFSEQ
Length	19282
Host Name	Bacillus subtilis
Host domain	Bacteria
Is prophage?	undef
Number of proteins	27
%GC	39.99
Number of singletons	1
Number of paralogs	0

List of proteins encoded in this genome

NCBI prot ID 11	PHROG prot ID	NCBI prot annotation	PHROG number	PHROG annotation	Strand 11	Start 1	End 1
YP_002004524.1	NC_011048_p1	hypothetical protein	singleton	unknown function	-1	265	360
YP_002004525.1	NC_011048_p2	hypothetical protein	phrog_38052	unknown function	-1	367	510
YP_002004526.1	NC_011048_p3	hypothetical protein	phrog_17724	unknown function	-1	527	703
YP_002004527.1	NC_011048_p4	hypothetical protein	phrog_13823	unknown function	-1	700	870
YP_002004528.1	NC_011048_p5	hypothetical protein	phrog_27272	unknown function	-1	857	1117
YP_002004529.1	NC_011048_p6	DNA polymerase	phrog_1907	DNA polymerase	-1	1145	2863
YP_002004530.1	NC_011048_p7	terminal protein	phrog_6358	DNA terminal protein	-1	2886	3686
YP_002004531.1	NC_011048_p8	transcriptional regulator	phrog_10365	late transcriptional activator	-1	3575	3952
YP_002004532.1	NC_011048_p9	single stranded DNA-binding protein	phrog_14995	single strand DNA binding protein	-1	3971	4345
YP_002004533.1	NC_011048_p10	hypothetical protein	phrog_26687	unknown function	-1	4410	4610
YP_002004534.1	NC_011048_p11	dsDNA binding protein	phrog_6370	dsDNA binding protein	-1	4607	4921
YP_002004535.1	NC_011048_p12	scaffolding protein	phrog_10439	head scaffolding protein	1	5222	5518
YP_002004536.1	NC_011048_p13	major head protein	phrog_2006	major head protein	1	5549	6895
YP_002004537.1	NC_011048_p14	head fiber protein	phrog_14753	head fiber protein	1	6895	7737
YP_002004538.1	NC_011048_p15	tail protein	phrog_2248	tail protein	1	7751	9550
YP_002004539.1	NC_011048_p16	upper collar protein	phrog_2038	upper collar connector	1	9555	10484
YP_002004540.1	NC_011048_p17	lower collar protein	phrog_3225	adaptor Ad4	1	10477	11358
YP_002004541.1	NC_011048_p18	preneck appendage protein	phrog_4775	pre-neck appendage protein	1	11371	13935

And summary tables with information on the genome and all of its encoded proteins

Genomic map of Bacillus virus phi29 (NC 011048)



You can click on proteins to see the summary page of their PHROG

More about this genome

Locus	NC_011048
Definition	Bacillus phage phi29, complete genome.
Molecule type	DNA
Genome structure	linear
Taxonomy	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Podoviridae; Picovirinae; Salasvirus.
DB of origin	REFSEQ
Length	19282
Host Name	Bacillus subtilis
Host domain	Bacteria
Is prophage?	undef
Number of proteins	27
%GC	39.99
Number of singletons	1
Number of paralogs	0

List of proteins encoded in this genome

NCBI prot ID 11	PHROG prot ID	NCBI prot annotation	PHROG number	PHROG annotation	Strand 11	Start 1	End 11
YP_002004524.1	NC_011048_p1	hypothetical protein	singleton	unknown function	-1	265	360
YP_002004525.1	NC_011048_p2	hypothetical protein	phrog_38052	unknown function	-1	367	510
YP_002004526.1	NC_011048_p3	hypothetical protein	phrog_17724	unknown function	-1	527	703
YP_002004527.1	NC_011048_p4	hypothetical protein	phrog_13823	unknown function	-1	700	870
YP_002004528.1	NC_011048_p5	hypothetical protein	phrog_27272	unknown function	-1	857	1117
YP_002004529.1	NC_011048_p6	DNA polymerase	phrog_1907	DNA polymerase	-1	1145	2863
YP_002004530.1	NC_011048_p7	terminal protein	phrog_6358	DNA terminal protein	-1	2886	3686
YP_002004531.1	NC_011048_p8	transcriptional regulator	phrog_10365	late transcriptional activator	-1	3575	3952
YP_002004532.1	NC_011048_p9	single stranded DNA-binding protein	phrog_14995	single strand DNA binding protein	-1	3971	4345
YP_002004533.1	NC_011048_p10	hypothetical protein	phrog_26687	unknown function	-1	4410	4610
YP_002004534.1	NC_011048_p11	dsDNA binding protein	phrog_6370	dsDNA binding protein	-1	4607	4921
YP_002004535.1	NC_011048_p12	scaffolding protein	phrog_10439	head scaffolding protein	1	5222	5518
YP_002004536.1	NC_011048_p13	major head protein	phrog_2006	major head protein	1	5549	6895
YP_002004537.1	NC_011048_p14	head fiber protein	phrog_14753	head fiber protein	1	6895	7737
YP_002004538.1	NC_011048_p15	tail protein	phrog_2248	tail protein	1	7751	9550
YP_002004539.1	NC_011048_p16	upper collar protein	phrog_2038	upper collar connector	1	9555	10484
YP_002004540.1	NC_011048_p17	lower collar protein	phrog_3225	adaptor Ad4	1	10477	11358
YP_002004541.1	NC_011048_p18	preneck appendage protein	phrog_4775	pre-neck appendage protein	1	11371	13935

You can zoom on the plot by scrolling or choosing a window to visualize

To get more information about a protein's PHROG, click on its respective link



III- Finding your protein in PHROGs

NOG: News Virus-Table Phrog-Table Suggestions Documentation



Welcome to the Prokaryotic Virus Remote Homologous Groups database (aka. PHROGS)

PHROG : families of prokaryotic virus proteins clustered using remote homology.

Terzian P*, Olo Ndela E*, Gallez C, Lossouam J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F.

This database contains 38,880 PHROGs (protein orthologous groups) containing 868,340 proteins from complete genomes of viruses infecting bacteria or archaea (2,318 from RetBeg and 2,669 from GenBank, april 2018), in addition to 12,498 curated prophages derived from cultivated microbial isolaties (Roux et al., 2015).

ninclosal socials (non-control of the control of th

This website provides access to :

 all prokanyotic virus genomes from the viruses table and select one to see its taxonomy, list of proteins, genomic map, etc...
 all PHROGs from the PHROGs table and select one to see its annotation, list of proteins, multiple alignment, comparison results to Pfam Unintrol. KEGG etc...

Viruses and PHROGs can also be access using search tools below.

insee and Fill consideration of access round search room bein

Find a PHROG of interest by searching in :

PHROG annotation terms	neck	Search
Refseq terms	neck	Search
Pfam domain number or annotation	PF05069.12	Search
KEGG Orthology or annotation	K14059	Search
GO term ID or annotation	lysozyme	Search

Or, if you already know the ID of your phrog you can go to its page directly here:

514

Find a virus of interest by searching in :

Virus Id NC_011048

If you don't know the accession of your virus, use the search box below :

Lambda

Virus name Find your protein :

PHROG number

If you have a protein from the NCBI and you wish to know if it belongs to a PHROG, you can search it by its sequence or its NCBI ID.

VCBI prot	D			
Sequence	search	(5	residues	minimu

NP_049616.1 Search

Go

Search

Search

Sequence search (5 residues minimum) MIRLLL Proteins that are not from the NCBI can also be retrieved through sequence search (PVOG, NUCCORE, and VirSorter).

imum) MIRLLL I can e search).

Download PHROGs as :

Fasta files Multiple alignments HMM profiles Annotation table

! Each of these zipped archive files (.tar.gz) contains 38,880 files (one fasta file, one MSA and HMM profile for each PHROG.

Out of the initial set of 938,864, a subset of 70,524 proteins were not grouped with other proteins into PHROG and remained singletons (or ORFans). You can also download these singleton protein as a single fasta file : Fasta files (Singletons)

Check out the manual on the Documentation page to compare your proteins to PHROGs

Virus and PHROG tables in tabular delimited format for downloading

PHROG table Virus Table



If you have a protein from a reference phage that was used in PHROGs' clustering, you can find it with this section of the home page

Find your protein :

If you have a protein from the NCBI and you wish to know if it belongs to a PHROG, you can search it by its sequence or its NCBI ID.



! Each of these zipped archive files (.tar.gz) contains 38,880 files (one fasta file, one MSA and HMM profile for each PHROG.

Out of the initial set of 938,864, a subset of 70,524 proteins were not grouped with other proteins into PHROG and remained singletons (or ORFans). You can also download these singleton protein as a single fasta file : Fasta files (Singletons)

Check out the manual on the Documentation page to compare your proteins to PHROGs.

Virus and PHROG tables in tabular delimited format for downloading :

PHROG table V

Virus Table

As a result you will have a table listing all the proteins with your ID, or sequence



Copy CSV Excel PDF

NCBI prot ID 1	NCBI prot annotation 11	PHROG prot	PHROG 1	PHROG annotation	↓ Virus ID ↑↓	Virus name	Virus taxonomy	¢↓
YP_002224718.1	gp20	NC_011272_p20	phrog_3040	no annotation	NC_011272	Mycobacterium phage Rizal	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus.	
AKY02318.1	hypothetical protein	KT321476_p24	phrog_3040	no annotation	KT321476	Mycobacterium phage Zeenon	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus; Mycobacterium virus Bxz1.	
YP_656033.1	gp20	NC_008207_p20	phrog_3040	no annotation	NC_008207	Mycobacterium phage Catera	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus.	
YP_009017359.1	gp23	NC_023733_p24	phrog_3040	no annotation	NC_023733	Mycobacterium phage MoMoMixon	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus.	
ALF51120.1	hypothetical protein	KT365400_p24	phrog_3040	no annotation	KT365400	Mycobacterium phage ErnieJ	Viruses; Duplodnavirla; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus; Mycobacterium virus Bxz1.	
ALF50885.1	hypothetical protein	KT365398_p22	phrog_3040	no annotation	KT365398	Mycobacterium phage DTDevon	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus; Mycobacterium virus Bxz1.	
AEJ94885.1	hypothetical protein	JF704096_p22	phrog_3040	no annotation	JF704096	Mycobacterium phage Ghost	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus; Mycobacterium virus Bxz1.	
NP_818099.1	gp23	NC_004687_p23	phrog_3040	no annotation	NC_004687	Mycobacterium phage Bxz1	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus.	
AVR57248.1	hypothetical protein	MH051266_p22	phrog_3040	no annotation	MH051266	Mycobacterium phage MikeLiesIn	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus; unclassified Bixzunavirus.	
YP_008060821.1	hypothetical protein	NC_021346_p24	phrog_3040	no annotation	NC_021346	Mycobacterium phage Gizmo	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus.	
AGV99741.1	hypothetical protein	KF024734_p21	phrog_3040	no annotation	KF024734	Mycobacterium phage Shrimp	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Bixzunavirus; Mycobacterium virus Bxz1.	

Search:



IV- Downloadable files

Find your protein :

If you have a protein from the NCBI and you wish to know if it belongs to a PHROG, you can search it by its sequence or its NCBI ID.



Comparing your sequence data to PHROGs : MMsegs

With **MMseqs** you can compare a single protein or a set of distinct unknown proteins to PHROGs. Start by downloading the PHROGs profile database for MMsegs : MMsegs PHROGs DB

(1) Untar the database and move to its directory to use it :

tar -xzf phrogs_mmseqs_db.tar.gz cd phrogs mmsegs db

(2) Create a database with your fasta file :

mmseqs createdb your_seq.faa target_seq

(3) Compute the search and convert the results into a tab separated file :

mmseqs search phrogs_profile_db target_seq results_mmseqs ./tmp mmseqs createtsv phrogs_profile_db target_seq results_mmseqs results.tsv

Comparing your sequence data to PHROGs : HHsuite

With HHsuite you can easily compare a single sequence, a multiple sequence alignment or a HMM profile to PHROGs using the following database and commands. If you have a multi-fasta file with a set of unknown proteins you'll have to build your custom index as specified below.

A good start is to download the indexed database here : HHsuite PHROGs DB

(1) Untar the database and move to its directory to use it :

tar -xzf phrogs_hhsuite_db.tar.gz cd phrogs hhsuite db

(2) Compute the search :

hhblits -i your_seq.faa -d phrogs -n 1 -o results_your_seq_VS_phrogs -blasttab tsv_file hhsearch -i your_seq.faa -d phrogs -n 1 -o results_your_seq_VS_phrogs -blasttab tsv_file

On the documentation page it is possible to download custom PHROGs databases for both MMseqs and HHsuite



V- Suggestions and News pages

PHROG annotations are flexible, therefore, you can suggest changes if you have expertise on a specific protein family



Welcome to the suggestions page !

Let's say you're working on your favorite virus and you have expertise on its gene functions, then you might want to precise or change the annotation of some PHROG(s).

If it is the case you're in the right place ! Use the form below to send us an email !

tour Email Address:
PHROG number:
Suggested annotation:
Tell us why:
Send

Annotation changes will be reported on the News page, as well as other important updates



Welcome to the news page !

Here you will be informed of every change in PHROGs since its first release ! The annotations of PHROGs are flexible, any noticed problem or suggested change can be reported in the Suggestions page.

Annotation release v2 - February 15 2021

A couple changes of annotation :

- phrog_9896 is now annotated as phosphofructokinase
- phrog_2185 is now annotated as RecT-like ssDNA annealing protein

These modifications are both based on PFAMs similar to these PHROGs.

Annotation v2

Annotation release v1

• This is the initial annotation release.

You can download the annotation of all PHROGs as a tab separeted file file here : Annotation v1

For more information, check out the Documentation page

0

PHROGS News Virus-Table Phrog-Table Suggestions Documentation

Any questions ?

Check out the Clustering pipeline and the user's guide below for more information. Please cite :

PHROG : families of prokaryotic virus proteins clustered using remote homology. Terzian P*, Olo Ndela E*, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit MA, Enault F.

Table of Contents

- Clustering and annotation pipeline
- User's guide
- Comparing your sequence data to PHROGs : MMseqs
- Comparing your sequence data to PHROGs : HHsuite







That's all folks !

Prokaryotic Virus Remote Homologous Groups (PHROG): clustering proteins from viruses of prokaryotes using remote homology detection.

Terzian Paul, Olo Ndela Eric, Galiez Clovis, Lossouarn Julien, Pérez Bucio Rubén Enrique, Mom Robin, Toussaint Ariane, Petit Marie-Agnès, Enault François

Report any bug to : phrogs.lmge@listes.uca.fr