



BLASTable Mycobacteriophage Database

Purpose: To have easy access to current known protein function through use of a DNA Master created protein database.

Background: When determining protein function, a variety of tools including BLASTP data, Phamerator maps, conserved domain identification, and HHpred are used. This is another tool to help. This database was created in DNA Master (procedure to follow) using 243 of the most recent mycobacteriophage genomes. This database was created December 2012.

Overview: This protocol consists of three parts:

- Creation of local database (supplied, see addendum for procedure)
- Import of database
- Local Blast of database

Procedure:

Import of Database

The library of completed mycobacteriophages with functions is currently a 3.41 MB file named Proteins_243MP_12.2012.dnamblib. This is a static file that will be periodically updated.

- Get Proteins_243MP_12.2012.dnamblib from the PhagesDB website. Click on Resources → Documents →Annotation Resources. Save this file to a known location.
- 2. Go to Tools → BLAST Library Manager



BLAST Library Manager	
Display All libraries 💌	Overview Groups Genomes Import/Export
Library Name	Use these functions to move BLAST libraries (though not genomes) between DNA Master installations
	Export Current BLAST Library
	Import DNA Master BLAST Library
Delete Select Genomes	
0 Import a DNA Master	format BLAST library

- 3. Click the Import/Export tab
- 4. Click the Import DNA Master BLAST Library button
- 5. Browse to the where the .dnamblib file is saved and select it.

You have now installed the Blast Library of proteins of 243 genomes. The next step is to blast your genome against these.

Local Blast of Database

 The genome must be added to the Genome Manager. To do this, open your genome in DNA Master. Click on Main menu→Genome→Add to Database. You can now close your genome file.

	ster			
File Edit	Genome DNA Tools Window Help	_		
	Add to Database			
Overview	Analyze all gene starts			
Genome	Auto-Annotate	axonomy No		
Organism	Bias Table	omain : Note		
GenomelE	BLAST All Genes	vision : Note		
Length:0	Coding Capacity	amily :		
Replicons Features	Codon Spacing	A		
CAI Table	Features >			
Directory :	Gene Orientation Ctrl+Alt+G			
Replicor	Karlin's Dinucleotides			
Replicon	Learn Features Ctrl+Alt+L			
NCBI Date	Mutational Bias			
Topology	Origin prediction			
Length : 5	Dendict stars	Note		

- 2. Now the sequence must be opened from the Genome Manager.
- 3. To open your file in Genome Manager: click on Tools \rightarrow Genome Manager

4. Double click on your genome to open the file. To blast this genome, it must be in the Genome Manager.



- 5. Now in this genome, go to the Feature tables and find a gene that has no blast data.
- 6. Click on the Blast tab and the following window appears.

🔀 Kykar_Draft					
Overview Features References	Sequence	Docume	entation		
Sort By Index 👻 🔳	Name	Start	Stop		Description Sequence Product Regions Blast Context
Select Features Direct SQL	▶ 1	597	893	_	There are no BLAST results for this feature
Turne in All all	2	928	1368		QBlast against public database via NCBI server
	3	1803	2873	Ξ	Blast this gene Blast ALL Genes Clear All
Name like	4	2883	3593		Local Blast against DNA Master database
GenelD =	5	3607	3867		Plast this game Plast All Games
Locus like	6	3875	4132		
Chud I	7	4132	5565		
	8	5555	6523		
Length <mark>></mark>	9	6543	8234	-	
Regions >	10	8234	9700		Click

- 7. Find Local Blast against DNA Master database and select the button Blast ALL Genes.
- 8. The Sequence Automation window pops up. Choose the following features.

DNA Sequence Automation			
Bias Tables Catenation Divergence Loc	cal Blast Orthologue Detectio	n Sequence Scan Skew	Input Sequences
Choose Local Blast Library			
243Genomes121312	▼	🔽 Is an External Library	Blast Manager
Save Results to Database			
🗖 Save In Local Database	Minimum Percent Identity	40.0	
✓ Clear previous results from this library	Minimum Percent Aligned	50.0	
	Maximum E-Value	0.0001	
Save Results To Text Files			
Save summary to file	Minimum Number per gene	1 单	
	Maximum Number per gene	10 🚖	
	Maximum E-Value	0.0001	
Perform Blast Cancel Blast			

- Choose the Local Blast library: 243Genomes121312.
- Under Save Results to Database: Use default settings.
- Under Save Results to Text Files: Checkmark the "Save summary to file" option. Use default settings for the remaining features.
- Click the Perform Blast button and choose where to save the output file.
- 9. The output file is a .csv file that can be opened in Excel.

	Δ	B	C		D	F	F	G	н	1	1		K		М	N	0
1	Idx	Ouervidx	TargetIdx	Ou	JervCoc	Ouerv	NOuervioc	TargetOrganism	Target@	TargetGei	r Targe	tloci	TargetProduct	Percentid	Alignmen I	PercentAl	EValue
2		1 1	1	59	7 - 893		1 DNAM1	Mycobacterium phage Solon	821	1	PBLS	010	Hypothetical Protein	98	98	98.99	1.20E-54
3		2 1	2	59	7 - 893		1 DNAM1	Mycobacterium phage KSSJEB	742	1	PBI K	SSIE	WhiB	95.9	98	98.99	3.50E-54
4		3 1	3	59	7 - 893		1 DNAM1	Mycobacterium phage MrGordo	762	1	PBIN	ARGC	HNH endonuclease	97.9	97	97.98	6.00E-54
5		4 1	4	59	7 - 893		1 DNAM1	Mycobacterium phage Doom	693	1	PBID	000	HNH endonuclease	96.9	98	98.99	6.00E-54
6		5 1	5	59	7 - 893		1 DNAM1	Mycobacterium phage Museum	764	3		NUSE	WhiB	96.9	98	98.99	1.30E-53
7		6 1	6	59	7 - 893		1 DNAM1	Mycobacterium phage JC27	732	2	PBIJ	C27	HNH-like domain	95.9	98	98.99	1.30E-53
8		7 1	7	59	7 - 893		1 DNAM1	PattyP ED	855	2		ATT	HNH endonuclease domain	95.9	98	98.99	1.70E-53
9		8 1	8	59	7 - 893		1 DNAM1	Mycobacterium phage Violet	843	2		/IOLE	HNH endonuclease domain	94.9	98	98.99	3.90E-53
10		9 1	9	59	7 - 893		1 DNAM1	Mycobacterium phage RidgeCB	803	2		RIDGE	HNH endonuclease domain	95.9	98	98.99	3.90E-53
11		0 1	10	59	7 - 893		1 DNAM1	Mycobacterium phage Perseus	781	1	PBI P	PERSE	HNH endonuclease domain	95.9	98	98.99	3.90E-53
12		1 2	2 1	92	8 - 1368		2 DNAM11	Mycobacterium phage Violet	843	3	B PBI V	/IOLE	Hypothetical Protein	100	146	99.32	3.20E-80
13		2 2	2 2	92	8 - 1368		2 DNAM11	Mycobacterium phage RidgeCB	803	3	PBI R	RIDGE	Hypothetical Protein	100	146	99.32	3.20E-80
14		3 2	2 3	92	8 - 1368		2 DNAM11	Mycobacterium phage Museum	764	4		NUSE	Hypothetical Protein	100	146	99.32	3.20E-80
15		4 2	2 4	92	8 - 1368		2 DNAM11	Mycobacterium phage MrGordo	762	3		MRGC	tail protein	100	146	99.32	3.20E-80
16		5 2	2 5	92	8 - 1368		2 DNAM11	Mycobacterium phage KSSJEB	742	2		SSJE	Hypothetical Protein	100	146	99.32	3.20E-80
17		.6 2	2 6	92	8 - 1368		2 DNAM11	Mycobacterium phage Doom	693	3	B PBI D		Hypothetical Protein	100	146	99.32	3.20E-80
18		7 2	2 7	92	8 - 1368		2 DNAM11	Mycobacterium phage Bxb1	667	2	PBI B	3XB1	related to L5 gp5 (41%); 16.3	100	146	99.32	3.20E-80
19		.8 2	2 8	92	8 - 1368		2 DNAM11	Mycobacterium phage Lesedi	747	2	2 PBI L	ESED	tail	99.3	146	99.32	1.20E-79
20		9 2	2 9	92	8 - 1368		2 DNAM11	Mycobacterium phage Lesedi	747	2	2 PBI L	ESED	tail	99.3	146	99.32	1.20E-79
21		20 2	2 10	92	8 - 1368		2 DNAM11	Mycobacterium phage Solon	821	2	PBI S	OLO	Hypothetical Protein	98.6	146	99.32	1.30E-78
22		1 3	3 1	18	03 - 287		3 DNAM21	Mycobacterium phage Perseus	781	6	5 PBI P	PERSE	Hypothetical Protein	98.3	356	99.72	0.00E+00
23		2 3	3 2	18	03 - 287		3 DNAM21	Mycobacterium phage Lockley	754	5	BI L	ockle	tail fiber-like protein	98	356	99.72	0.00E+00
24		3 3	3	18	03 - 287		3 DNAM21	Mycobacterium_phage_Lesedi	747	4	PBI_L	ESED	Hypothetical_Protein	98	356	99.72	0.00E+00
25		4 3	3 4	18	03 - 287		3 DNAM21	Mycobacterium phage Lesedi	747	4	PBI L	ESED	Hypothetical Protein	98	356	99.72	0.00E+00
26		25 3	5 5	18	03 - 287		3 DNAM21	Mycobacterium phage JC27	732	6	5 PBI J	C27	tail	97.5	356	99.72	0.00E+00
14 4	()) I	vkar Draft	/ *] /														

Notes:

- This is a static database created December 2012 with 243 mycobacteriophage genomes.
- At present, you can only retrieve the best 10 hits.
- Functions are present in this database as called by Phage Hunters. Information obtained should still undergo evaluation.

Addendum:

Creation of Local Database

- 1. To create a local database, go to Tools \rightarrow Genome Manager.
- 2. Select the genomes you wish to place in the library by importing desired genomes into your version of DNA Master Genome Manager. Any genome with a DNA Master file or one found in GenBank can be added. To add a DNA Master file, see first step of Local Blast of Database. To add a file from GenBank, go to the Retrieve tab in Genome Manager. From the submenu there, select the Fetch by Accession tab. If you know the accession number(s), place that number(s) in the window on the left side of the window. Once your list is complete, click Save in local database.

Manager Tools Window Help		
Genome Manager		
rowse Search Taxonomy Retrieve rRNA	Genes Favorites and Clipboard Tools	
Complete Genome RefSeq Entry Fetch By Ac	cession Partial Genome	
Fetch by Accession Search NCB	I for	Show 100
C Open sequences ✓ Auto-save files ✓ Auto-open files ✓ Save in local database	Keep in Mind Open	Found : 0
Tag DWN		
Separate Genomes		
0398041		
A000041		
*		
•		

3. Click on the Browse tab. This will populate your Genome Manager with as many genomes as you choose.

DNA Master ile Manager Tools Window Help				
Genome Manager				
Browse Search Taxonomy Retrieve RNA Genes F	avorites and C	ipboard Tools		
Organism 🖉 🖌	Genome D	escription Organis	sm Taxonomy Tags, I	Notes etc.
Mycobacterium phage 244 Mycobacterium phage ABU	Data Direc	tory DM500276	NCBI Directory	
Mycobacterium phage Acadian	Genome D	ate 1/16/2013	🔽 Features Le	amed
Mycobacterium phage Adephagia Mycobacterium phage Adjutor	Date Creat	ed 1/16/2013	CAI Table	•
Mycobacterium phage Airmid Mucobacterium phage Airmid	Total Featu Average %	ures 86 GC 00.00	Genome Length GenomelD	276
Mycobacterium phage Alice	Replicons	1	rRNA prototype	0 33
Mycobacterium phage Anaya	rRNAs Beplicon D	0	User TaxonID	aan Nataa l
Mycobacterium phage Angel Mycobacterium phage Angelica	Replicon	Replicon0	on breakpoints [riepin	
Mycobacterium phage Arbiter Mycobacterium phage Ardmore	RepliconID	0	Translation Table	Unspecified: default to standard
-	Date	1/9/2013	GI	-41200420050
Replicon Name Replicon0	% GC	00.00	Accession Version	x41290438850
	Features	86		🗖 Is circular
	ORFs	85		Is the primary replicon Is a catenated replicon
Open Genome Delete Genome Delete Replicon	111445	1		Is replaced by a more recent version Is protected and cannot be delete
227 Genomes 222 archaea Live List of geno	omes in local d	latabase		

4. Go to Favorites and Clipboard tab and add all genomes of interest to the clipboard (middle window).



- 5. Go to the Tools tab → Blast Library
- 6. Enter a name of the library, choose a protein library, specify that the Products and Organisms by included in the BLAST header.
 - DNA Master <u>File Manager Tools Window H</u>elp Genome Manager Browse Search Taxonomy Retrieve RNA Genes Favorites and Clipboard Tools Blast Library Log Organism Library Name <Enter Library Name here> Kykar Draft E Sequence Format
 Protein Library
 C Nucleotide Library Mycobacterium phage 244
 Mycobacterium phage ABU Included Features • Mycobacterium phage Acadian Mycobacterium phage Adephagia 🔽 Include All Replicons Mycobacterium phage Adjutor Mycobacterium phage Airmid ☑ Include Product in BLAST Header ✓ Include Organism in BLAST Header Mycobacterium phage Akoma Create Blast Library Notes and Comments Mycobacterium phage Alice Mycobacterium phage Alice Mycobacterium phage Anaya Mycobacterium phage Angel Mycobacterium phage Angelica Mycobacterium phage Arbiter Mycobacterium phage Ardmore Mycobacterium phage Ares Mycobacterium phage Athena Mycobacterium phage Avrafan Mycobacterium phage BPBiebs31 Mycobacterium phage BPs Mycobacterium phage Babsiella Mycobacterium phage Backyardigan 227 Genomes 1190 on NCBI Live
- 7. Create the library
- 8. This will create a .dnamblib file with the name you chose in step 5.
- 9. Proceed to Import of Database.