Poxvirus Bioinformatics Resource Center: a comprehensive *Poxviridae* informational and analytical resource

Elliot J. Lefkowitz*, Chris Upton[1], Shankar S. Changayil, Charles Buck[2], Paula Traktman[3] and R. Mark L. Buller[4]

Department of Microbiology, University of Alabama at Birmingham, BBRB 276/11; 1530 3rd Avenue S., Birmingham, AL 35294-2170, USA, [1]Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada V8W 2Y2, [2]Virology Collection, ATCC, Manassas, VA 20108, USA, [3]Department of Microbiology and Molecular Genetics, Medical College of Wisconsin, Room 273–BSB, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA and [4]Department of Molecular Microbiology and Immunology, St Louis University Health Sciences Center, 1402 South Grand Boulevard, St Louis, MO 63104, USA

Received August 12, 2004; Revised and Accepted October 18, 2004

ABSTRACT

The Poxvirus Bioinformatics Resource Center (PBRC) has been established to provide informational and analytical resources to the scientific community to aid research directed at providing a better understanding of the *Poxviridae* family of viruses. The PBRC was specifically established as the result of the concern that variola virus, the causative agent of smallpox, as well as related viruses, might be utilized as biological weapons. In addition, the PBRC supports research on poxviruses that might be considered new and emerging infectious agents such as monkeypox virus. The PBRC consists of a relational database and web application that supports the data storage, annotation, analysis and information exchange goals of the project. The current release consists of over 35 complete genomic sequences of various genera, species and strains of viruses from the *Poxviridae* family. Sequence and annotation information for these viruses has been obtained from sequences publicly available from GenBank as well as sequences not yet deposited in GenBank that have been obtained from ongoing sequencing projects. In addition to sequence data, the PBRC provides comprehensive annotation and curation of virus genes; analytical tools to aid in the understanding of the available sequence data, including tools for the comparative analysis of different virus isolates; and visualization tools to help better display the results of various analyses. The PBRC represents the initial development of what will become a more comprehensive Viral Bioinformatics Resource Center for Biodefense that will be one of the National Institute of Allergy and Infectious Diseases' 'Bioinformatics Resource Centers for Biodefense and Emerging or Re-Emerging Infectious Diseases'. The PBRC website is available at http://www.poxvirus.org.

INTRODUCTION

An effective response to the use of biological organisms as agents of terrorism or warfare, or to the emergence of new infectious diseases requires a multi-disciplinary effort involving various agencies at the local, state and federal levels including public health officials, hospital personnel, epidemiologists and the military. In addition to the public health response, a concerted research effort is necessary to better detect, understand and respond to these threats. Such research requires development of environmental detectors and clinical diagnostic aids to provide us with rapid warning in the event of an outbreak as well as development of vaccines to prevent infection and antiviral or antibacterial drugs to cure infection. These efforts require a comprehensive biological understanding of potential threat agents, including their molecular biology, genetics, pathogenicity, epidemiology and evolution. The National Institute of Allergy and Infectious Diseases as well as the US Centers for Disease Control and Prevention maintain a list of priority pathogens that are considered potential biothreat agents and/or are microbes that appear to be new or

*To whom correspondence should be addressed: Tel: +1 205 934 1946; Fax: +1 205 934 9256; Email: elliotl@uab.edu reemerging pathogens (http://www.bt.cdc.gov/agent/agentlist-category.asp and http://www.niaid.nih.gov/biodefense/bandc_priority.htm). Variola virus, the causative agent of smallpox and a member of the *Poxviridae* family of viruses, has perhaps the greatest potential for use as a bio-weapon and is one of the Category A pathogens on these priority pathogen lists (1). In addition, monkeypox virus, a member of the orthopoxvirus genus that includes variola virus, has caused a number of disease outbreaks in recent years, including outbreaks in North America resulting from the importation of rodents from Africa intended to be sold as pets (2,3).

The use of high-throughput DNA sequencing techniques as well as other large-scale 'Systems Biology' technologies have led to an unprecedented increase in the amount of available data. Therefore, one overarching necessity in research efforts directed at providing a better understanding of priority pathogens is the need to collect, manage, describe, analyze and publicize the vast amounts of information generated by modern, high-throughput biological research. Therefore, the goal of the Poxvirus Bioinformatics Resource Center (PBRC) is to organize all available information on virus genetics thus aiding research efforts towards increasing our knowledge of virus replication and virus–host interaction on a gene-by-gene and whole genome basis. In addition, the PBRC is expanding on available knowledge by developing and utilizing analysis tools that can further probe the information contained in the genome and gene sequences of these organisms. Since our goal is to establish an information resource to support research efforts by the scientific community, we are also soliciting input from that community to ensure the completeness and, above all, the accuracy of the information being provided and to ensure that the software tools provided and in development reflect the needs of the different research groups using these resources.

WHAT IS THE POXVIRUS BIOINFORMATICS RESOURCE CENTER?

The PBRC represents a cooperative endeavor with collaborations between the University of Alabama at Birmingham, the University of Victoria, the Medical College of Wisconsin, St Louis University and the American Type Culture Collection (ATCC). The PBRC consortium has established a database of all available completely sequenced poxvirus genomes. This database includes information on every predicted gene that may be coded for by these genomes, as well as descriptive annotations of the physical and functional properties of each gene based on computer predictions. Currently, only complete genomes are included in the database. In future releases we plan to include as available, incomplete genomes as well as coding sequences from virus strains not represented by complete genomic sequences. We are also compiling a comprehensive gene-by-gene curation that is linked to each individual gene record. In addition, the PBRC provides a variety of analytical information and analysis tools that can be used to mine the genomic data. These tools include sequence homology searches, a database of functional domains, a database of poxvirus gene orthologs and web-based visualization tools to allow for customized displays of much of this information. A major goal has been to develop new software packages for the analysis of viral genomes. This aspect of the project involves designing new software tools that permit researchers to interact with and manipulate complete poxvirus genomes and families of poxvirus protein orthologs. This vastly speeds up the analysis process and provides information about these viruses from comparative analyses that otherwise would be almost impossible to obtain. Some of these tools have been described previously in the literature and therefore more substantial descriptions are already available [Poxvirus Orthologous Clusters (4,5), Viral Genome Organizer (6), Viral Genome Database (7), JDotter (8) and Base-By-Base (9)].

DATABASE DESCRIPTION

The basic PBRC web portal accesses genomic, annotative and analytical information from a Microsoft SQL Server database. A companion MySQL database provides data to a number of java-based analytical tools. The database schema used for the PBRC originally was developed to accommodate bacterial genomes and their genes (protein and RNA) (10,11) and needed only slight modification to support the storage of information on poxvirus genomes and genes (12). The PBRC data schema provides tables to store sequences and basic sequence annotations, human-annotated curation records and tables for the storage of basic analytical information such as the results of BLAST searches, functional motifs and biophysical properties. The database supports all of the web-based query tools and also serves as the data source for all of our web-based analytical tools.

GENOME DATA ACQUISITION

Genomic data are obtained from GenBank (13), other publicly available databases and websites, as well as from ongoing sequencing projects. The current PBRC release consists of over 35 complete genomes from the *Poxviridae* family of viruses. If available from the GenBank record or from the sequencing laboratory, the predicted gene set of any one genome is stored in the PBRC database and used as the initial starting point for annotation and curation of each genome. If a predicted gene set is not available, we perform our own gene prediction using a combination of gene predictive tools that start with open reading frames, and then include among others, promoter prediction, presence of functional motifs and assignment to orthologous protein clusters. Our experience is that the methods used for the prediction of protein coding genes for each poxvirus genome present in GenBank has in many cases been derived using different parameters. Therefore, we are in the process of developing a more consistent method for the prediction of poxvirus genes and will utilize this gene prediction pipeline to reassess the gene set for each available genome.

Each virus genome is categorized according to its genus, species and strain designation as determined by the International Committee for the Taxonomy of Viruses (ICTV) (http://www.ncbi.nlm.nih.gov/ICTVdb/) (14,15). In addition to providing the nomenclature available from the GenBank record for each gene in any one genome, we also provide our own gene designation which is a combination of the ICTV-approved species name, the strain or isolate name and a numerical designation for each gene that starts at number 1 for the left-most gene, and is incremented by one for each subsequent gene as determined by its genome position.

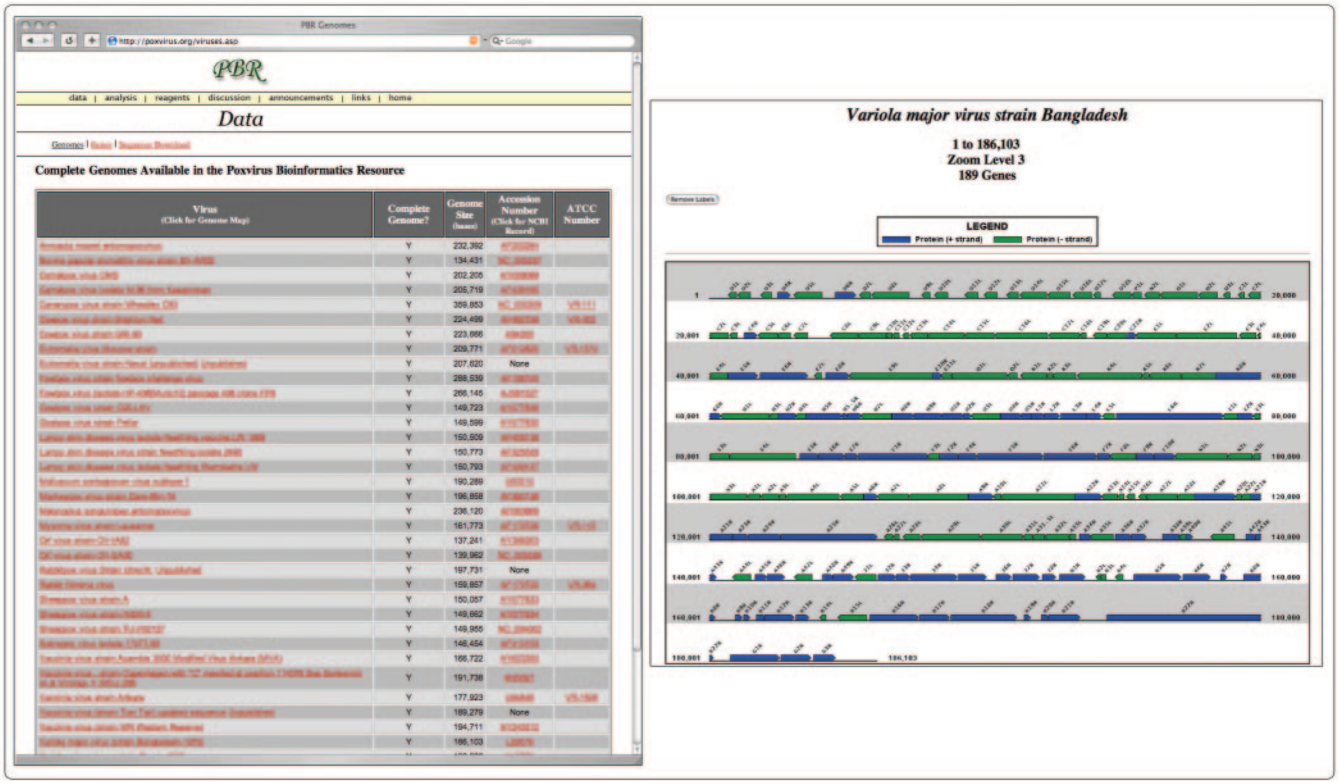
GENOME ANNOTATION AND CURATION

For each gene in the PBRC database, we provide an automated, computer-driven annotation that gathers as much basic descriptive information as possible about a gene, basic analysis of its nucleotide or amino acid sequence, and the results of sequence similarity searches to look for common patterns or features that might be characteristic of its function. The annotation process starts with the GenBank record and includes the descriptive information, literature references and any other information provided in that record. This information populates the initial descriptive fields of our database. Following this automated annotation process, a manual, human-directed curation of each gene record is undertaken. During this curation process, a researcher reviews the annotation record, all available literature references and any unpublished information as available. This collection of empirically derived properties for the protein in question provides what might be considered a mini-review of the biology of the gene being studied. The broad types of information that are provided during the curation process include protein properties such as molecular weight and pI; post-translational processing; the availability of custom reagents such as clones, antibodies and mutants; functional descriptions [including Gene Ontology designations (16)]; and literature summaries. Evidence codes are provided that explicitly state the nature and source of each piece of information along with the appropriate literature references. A series of web forms assist in this process that provides a distinct set of informational fields to be filled in, and enforces use of a controlled vocabulary to fully describe each gene. The results of the curation process are stored in our SQL Server database and form a Poxvirus Knowledge Database that is available and searchable from the PBRC website.

WEBSITE

The PBRC website is provided by a Microsoft Windows 2003 server running Microsoft Internet Information Services (IIS). The user interface is provided through a combination of Active Server Pages running server-side Visual Basic script, client-side JavaScript and HTML. SQL Server database access is provided through Microsoft ActiveX Data Objects.

The starting point for access to all databases and tools available from the PBRC is the PBRC home page at http://www.poxvirus.org/. A menu is displayed along the top of the page and appropriate clicking on a menu link brings up a submenu that provides access to individual web query tools and/or applications. The Data menu provides access to search forms that provide user-specified queries for genome, gene and sequence data (Figures 1 and 2). Analytical tools are provided through another series of submenus and provide access to a series of analytical and visualization tools (Figure 3). In addition to the main menu and context-sensitive


PBR Genomes
http://poxvirus.org/viruses.asp
Google
PBR
data | analysis | reagents | discussion | announcements | links | home
Data
Genomes | [illegible] | [illegible]

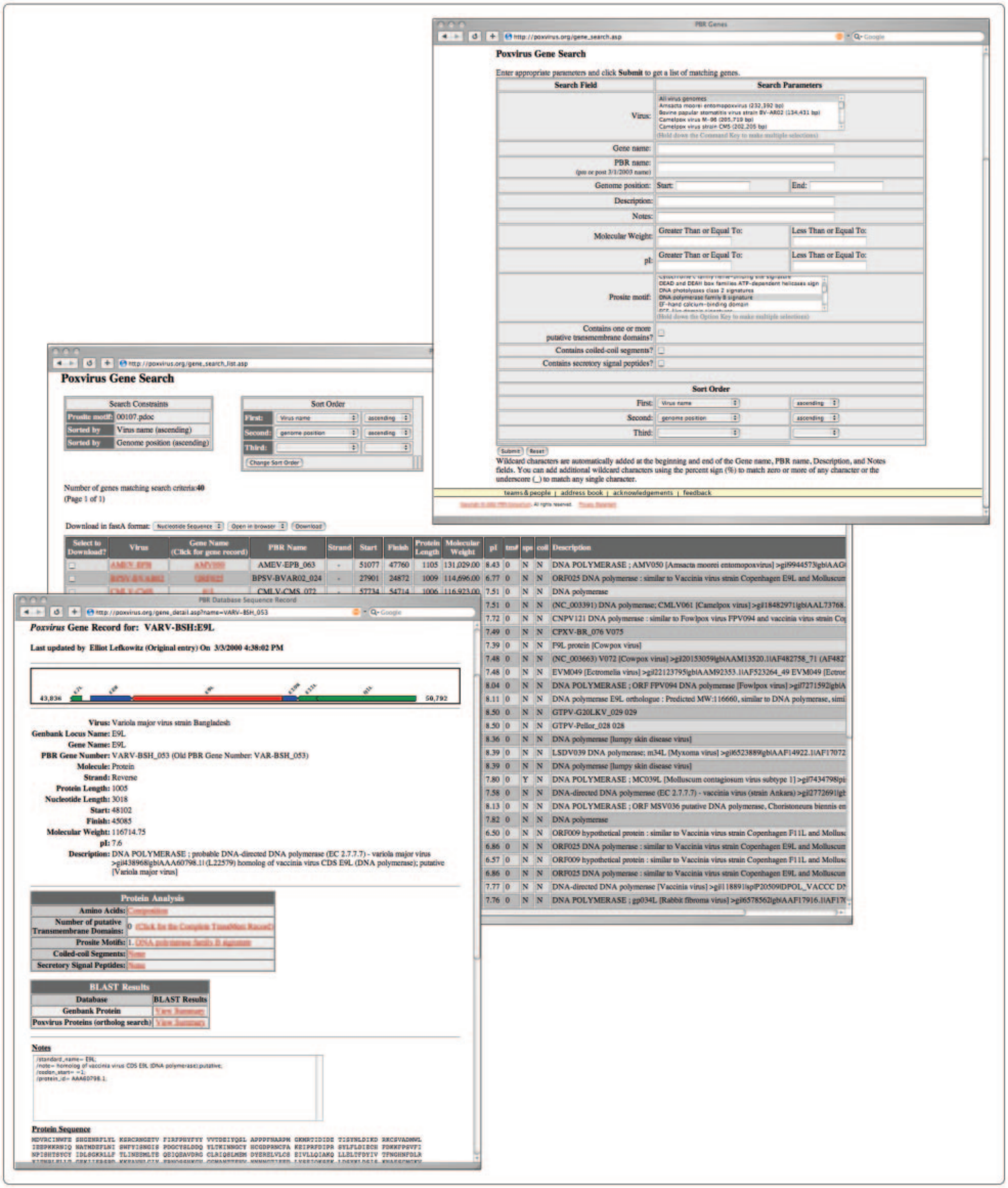
Complete Genomes Available in the Poxvirus Bioinformatics Resource

Virus (Click for Genome Map)	Complete Genome?	Genome Size (bases)	Accession Number (Click for NCBI Record)	ATCC Number
[illegible]	Y	232,392	[illegible]	
[illegible]	Y	134,431	[illegible]	
[illegible]	Y	202,205	[illegible]	
[illegible]	Y	205,719	[illegible]	
[illegible]	Y	359,853	[illegible]	[illegible]
[illegible]	Y	224,499	[illegible]	[illegible]
[illegible]	Y	223,666	[illegible]	
[illegible]	Y	209,771	[illegible]	[illegible]
[illegible]	Y	207,620	None	
[illegible]	Y	288,539	[illegible]	
[illegible]	Y	266,145	[illegible]	
[illegible]	Y	149,723	[illegible]	
[illegible]	Y	149,599	[illegible]	
[illegible]	Y	150,509	[illegible]	
[illegible]	Y	150,773	[illegible]	
[illegible]	Y	150,793	[illegible]	
[illegible]	Y	190,289	[illegible]	
[illegible]	Y	196,858	[illegible]	
[illegible]	Y	236,120	[illegible]	
[illegible]	Y	161,773	[illegible]	[illegible]
[illegible]	Y	137,241	[illegible]	
[illegible]	Y	139,962	[illegible]	
[illegible]	Y	197,731	None	
[illegible]	Y	159,857	[illegible]	[illegible]
[illegible]	Y	150,057	[illegible]	
[illegible]	Y	149,662	[illegible]	
[illegible]	Y	149,955	[illegible]	
[illegible]	Y	146,454	[illegible]	
[illegible]	Y	166,722	[illegible]	
[illegible]	Y	191,738	[illegible]	
[illegible]	Y	177,923	[illegible]	[illegible]
[illegible]	Y	189,279	None	
[illegible]	Y	194,711	[illegible]	
[illegible]	Y	186,103	[illegible]	

Variola major virus strain Bangladesh

1 to 186,103
Zoom Level 3
189 Genes

Figure 1. PBRC genomic sequence web pages. Screen shots of the PBRC genome list, and the genome map for *Variola major* virus strain Bangladesh.



Poxvirus Gene Search

Enter appropriate parameters and click **Submit** to get a list of matching genes.

Search Field	Search Parameters	
Virus:	All virus genomes Amsacta moorei entomopoxvirus (232,392 bp) Bovine papular stomatitis virus strain BV-AR02 (134,431 bp) Camelpox virus M-96 (205,719 bp) Camelpox virus strain CMS (202,205 bp) (Hold down the Command Key to make multiple selections)	
Gene name:		
PBR name: (pre or post 3/1/2003 name)		
Genome position:	Start:	End:
Description:		
Notes:		
Molecular Weight:	Greater Than or Equal To:	Less Than or Equal To:
pI:	Greater Than or Equal To:	Less Than or Equal To:
Prosite motif:	DEAD and DEAH box families ATP-dependent helicases sign DNA photolyases class 2 signatures DNA polymerase family B signature EF-hand calcium-binding domain (Hold down the Option Key to make multiple selections)	
Contains one or more putative transmembrane domains?	□	
Contains coiled-coil segments?	□	
Contains secretory signal peptides?	□	
	Sort Order	
First:	Virus name	ascending
Second:	genome position	ascending
Third:		

Wildcard characters are automatically added at the beginning and end of the Gene name, PBR name, Description, and Notes fields. You can add additional wildcard characters using the percent sign (%) to match zero or more of any character or the underscore (_) to match any single character.

teams & people | address book | acknowledgements | feedback
All rights reserved.
http://poxvirus.org/gene_search_list.asp

Poxvirus Gene Search

Search Constraints	
Prosite motif:	00107.pdoc
Sorted by	Virus name (ascending)
Sorted by	Genome position (ascending)

Sort Order		
First:	Virus name	ascending
Second:	genome position	ascending
Third:		

Number of genes matching search criteria:**40**
(Page 1 of 1)

Download in fastA format: Nucleotide Sequence | Open in browser | Download

Select to Download?	Virus	Gene Name (Click for gene record)	PBR Name	Strand	Start	Finish	Protein Length	Molecular Weight	pI	tm#	sps	coil	Description
□	AMEV-EPB	AMV050	AMEV-EPB_063	-	51077	47760	1105	131,029.00	8.43	0	N	N	DNA POLYMERASE ; AMV050 [Amsacta moorei entomopoxvirus] >gi|9944573|gb|AAG
□	BPSV-BVAR02	ORF025	BPSV-BVAR02_024	-	27901	24872	1009	114,696.00	6.77	0	N	N	ORF025 DNA polymerase : similar to Vaccinia virus strain Copenhagen E9L and Molluscum
□	CMLV-CMS	[illegible]	CMLV-CMS_072	-	57734	54714	1006	116,923.00	7.51	0	N	N	DNA polymerase
									7.51	0	N	N	(NC_003391) DNA polymerase; CMLV061 [Camelpox virus] >gi|18482971|gb|AAL73768.
									7.72	0	N	N	CNPV121 DNA polymerase : similar to Fowlpox virus FPV094 and vaccinia virus strain Co
									7.49	0	N	N	CPXV-BR_076 V075
									7.39	0	N	N	F9L protein [Cowpox virus]
									7.48	0	N	N	(NC_003663) V072 [Cowpox virus] >gi|20153059|gb|AAM13520.1|AF482758_71 (AF482
									7.48	0	N	N	EVM049 [Ectromelia virus] >gi|22123795|gb|AAM92353.1|AF523264_49 EVM049 [Ectro
									8.04	0	N	N	DNA POLYMERASE ; ORF FPV094 DNA polymerase [Fowlpox virus] >gi|7271592|gb|A
									8.11	0	N	N	DNA polymerase E9L orthologue : Predicted MW:116660, similar to DNA polymerase, simi
									8.50	0	N	N	GTPV-G20LKV_029 029
									8.50	0	N	N	GTPV-Pellor_028 028
									8.36	0	N	N	DNA polymerase [lumpy skin disease virus]
									8.39	0	N	N	LSDV039 DNA polymerase; m34L [Myxoma virus] >gi|6523889|gb|AAF14922.1|AF17072
									8.39	0	N	N	DNA polymerase [lumpy skin disease virus]
									7.80	0	Y	N	DNA POLYMERASE ; MC039L [Molluscum contagiosum virus subtype 1] >gi|7434798|pi
									7.58	0	N	N	DNA-directed DNA polymerase (EC 2.7.7.7) - vaccinia virus (strain Ankara) >gi|2772691|gb
									8.13	0	N	N	DNA POLYMERASE ; ORF MSV036 putative DNA polymerase, Choristoneura biennis en
									7.82	0	N	N	DNA polymerase
									6.50	0	N	N	ORF009 hypothetical protein : similar to Vaccinia virus strain Copenhagen F11L and Mollusc
									6.86	0	N	N	ORF025 DNA polymerase : similar to Vaccinia virus strain Copenhagen E9L and Molluscum
									6.57	0	N	N	ORF009 hypothetical protein : similar to Vaccinia virus strain Copenhagen F11L and Mollusc
									6.86	0	N	N	ORF025 DNA polymerase : similar to Vaccinia virus strain Copenhagen E9L and Molluscum
									7.77	0	N	N	DNA-directed DNA polymerase [Vaccinia virus] >gi|118891|sp|P20509|DPOL_VACCC DN
									7.76	0	N	N	DNA POLYMERASE ; gp034L [Rabbit fibroma virus] >gi|6578562|gb|AAF17916.1|AF170

PBR Database Sequence Record
http://poxvirus.org/gene_detail.asp?name=VARV-BSH_053
Google

***Poxvirus* Gene Record for: VARV-BSH:E9L**

Last updated by Elliot Lefkowitz (Original entry) On 3/3/2000 4:38:02 PM

Virus: Variola major virus strain Bangladesh
Genbank Locus Name: E9L
Gene Name: E9L
PBR Gene Number: VARV-BSH_053 (Old PBR Gene Number: VAR-BSH_053)
Molecule: Protein
Strand: Reverse
Protein Length: 1005
Nucleotide Length: 3018
Start: 48102
Finish: 45085
Molecular Weight: 116714.75
pI: 7.6
Description: DNA POLYMERASE ; probable DNA-directed DNA polymerase (EC 2.7.7.7) - variola major virus >gi\|438968\|gb\|AAA60798.1\| (L22579) homolog of vaccinia virus CDS E9L (DNA polymerase); putative [Variola major virus]

Protein Analysis	
Amino Acids:	Composition
Number of putative Transmembrane Domains:	0 (Click for the Complete TransMem Record)
Prosite Motifs:	1. DNA polymerase family B signature
Coiled-coil Segments:	None
Secretory Signal Peptides:	None

BLAST Results	
Database	BLAST Results
Genbank Protein	View Summary
Poxvirus Proteins (ortholog search)	View Summary

Notes

```
/standard_name= E9L;
/note= homolog of vaccinia virus CDS E9L (DNA polymerase);putative;
/codon_start= =1;
/protein_id= AAA60798.1;
```

Protein Sequence

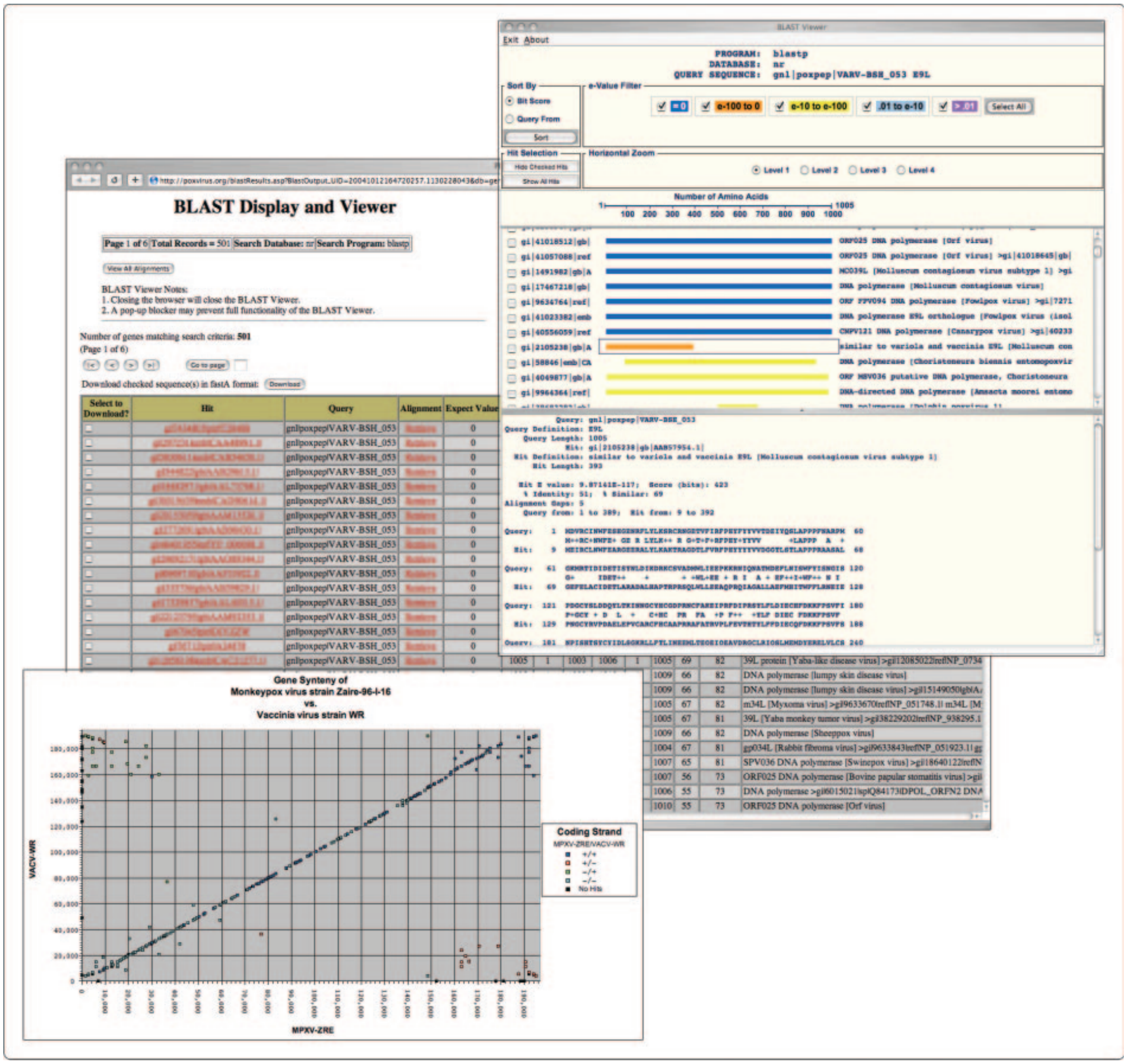
```
MDVRCINWFE SHGENRFLYL KSRCRNGETV FIRFPHYFYY VVTDEIYQSL APPPFNARPM GKMRTIDIDE TISYNLDIKD RKCSVADMWL
IEEPKKRNIQ NATMDEFLNI SWFYISNGIS PDGCYSLDDQ YLTKINNGCY HCDDPRNCFA KEIPRFDIPR SYLFLDIECH FDKKFPSVFI
NPISHTSYCY IDLSGKRLLF TLINEEMLTE QEIQEAVDRG CLRIQSLMEM DYERELVLCS EIVLLQIAKQ LLELTFDYIV TFNGHNFDLR
```

Figure 2. PBRC gene search and gene record web pages. Screen shots of the PBRC gene search form, a listing of gene search results and an individual gene record.

submenu, a menu near the bottom of the page provides access to PBRC organizational information such as descriptions of the people and teams involved in the work, acknowledgements and a form to provide user feedback. At the very bottom of the form we provide a Google-supported search form.

In general, available web pages can be categorized as follows:

(i) *Informational*: Static text such as help files, biographical information or meeting announcements.
(ii) *Forms*: Pages that provide user input for data queries.



```
PROGRAM:  blastp
DATABASE:  nr
QUERY SEQUENCE:  gnl|poxpep|VARV-BSH_053 E9L
```

```
☐ gi|41018512|gb|    ORF025 DNA polymerase [Orf virus]
☐ gi|41057088|ref    ORF025 DNA polymerase [Orf virus] >gi|41018645|gb|
☐ gi|1491982|gb|A    MC039L [Molluscum contagiosum virus subtype 1] >gi
☐ gi|17467218|gb|    DNA polymerase [Molluscum contagiosum virus]
☐ gi|9634764|ref|    ORF FPV094 DNA polymerase [Fowlpox virus] >gi|7271
☐ gi|41023382|emb    DNA polymerase E9L orthologue [Fowlpox virus (isol
☐ gi|40556059|ref    CNPV121 DNA polymerase [Canarypox virus] >gi|40233
☐ gi|2105238|gb|A    similar to variola and vaccinia E9L [Molluscum con
☐ gi|58846|emb|CA    DNA polymerase [Choristoneura biennis entomopoxvir
☐ gi|4049877|gb|A    ORF MSV036 putative DNA polymerase, Choristoneura
☐ gi|9964364|ref|    DNA-directed DNA polymerase [Amsacta moorei entomo
```

```
           Query: gnl|poxpep|VARV-BSE_053
Query Definition: E9L
    Query Length: 1005
             Hit: gi|2105238|gb|AAB57954.1|
  Hit Definition: similar to variola and vaccinia E9L [Molluscum contagiosum virus subtype 1]
      Hit Length: 393

   Hit E value: 9.87141E-117;  Score (bits): 423
    % Identity: 51;  % Similar: 69
Alignment Gaps: 5
    Query from: 1 to 389;  Hit from: 9 to 392

Query:    1  MDVRCINWFESHGENRFLYLKSRCRNGETVFIRFPHYFYYVVTDEIYQSLAPPPFNARPM  60
             M++RC+NWFE+ GE R LYLK++ R G+T+F+RFPHY+YYVV       +LAPPP  A  +
  Hit:    9  MEIRCLNWFEARGESRALYLKAKTRAGDTLFVRFPHYYYYVVDGGTLSTLAPPPRAASAL  68

Query:   61  GKMRTIDIDETISYNLDIKDRKCSVADMWLIEEPKKRNIQNATMDEFLNISWFYISNGIS 120
             G+      IDET++    +         + +WL+EE + R I  A + EF++I+WF+ N I
  Hit:   69  GEFELACIDETLARADALHAPTRPRSQLWLLEEAQPRQIAGALLAEFMHITWFFLANEIE 128

Query:  121  PDGCYSLDDQYLTKINNGCYHCGDPRNCFAKEIPRFDIPRSYLFLDIECHFDKKFPSVFI 180
             P+GCY + D  L  +    C+HC  PR  FA  +P F++  +YLF DIEC FDKKFPSVF
  Hit:  129  PNGCYRVPDAELEPVCARCPHCAAPRRAFATRVPLFEVTRTYLFFDIECQFDKKFPSVFS 188

Query:  181  NPISHTSYCYIDLSGKRLLFTLINEEMLTEDEIQEAVDRGCLRIQSLMEMDYERELVLCS 240
```

BLAST Display and Viewer

Page 1 of 6	Total Records = 501	Search Database: nr	Search Program: blastp

BLAST Viewer Notes:
1. Closing the browser will close the BLAST Viewer.
2. A pop-up blocker may prevent full functionality of the BLAST Viewer.

Number of genes matching search criteria: **501**
(Page 1 of 6)

Download checked sequence(s) in fastA format:

Select to Download?	Hit	Query	Alignment	Expect Value
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0
☐	[illegible]	gnl|poxpep|VARV-BSH_053	[illegible]	0

1005	1	1003	1006	1	1005	69	82	39L protein [Yaba-like disease virus] >gi|12085022|ref|NP_0734
					1009	66	82	DNA polymerase [lumpy skin disease virus]
					1009	66	82	DNA polymerase [lumpy skin disease virus] >gi|15149050|gb|A
					1005	67	82	m34L [Myxoma virus] >gi|9633670|ref|NP_051748.1| m34L [M
					1005	67	81	39L [Yaba monkey tumor virus] >gi|38229202|ref|NP_938295.1
					1009	66	82	DNA polymerase [Sheeppox virus]
					1004	67	81	gp034L [Rabbit fibroma virus] >gi|9633843|ref|NP_051923.1| gp
					1007	65	81	SPV036 DNA polymerase [Swinepox virus] >gi|18640122|ref|N
					1007	56	73	ORF025 DNA polymerase [Bovine papular stomatitis virus] >gi
					1006	55	73	DNA polymerase >gi|6015021|sp|Q84173|DPOL_ORFN2 DNA
					1010	55	73	ORF025 DNA polymerase [Orf virus]

Figure 3. PBRC BLAST search and gene synteny analytical tool web pages. Screen shots of XS-BLAST tabular and graphical BLAST search results and a gene synteny plot.

(iii) *Tabular results*: Pages displaying data in tables.
(iv) *Graphical results*: Pages displaying data results as figures.
(v) *Links*: Pages providing access to other pages.
(vi) *Applets*: Small Java-based applications for data analysis and visualization that are available from within a user's web browser.
(vii) *Applications*: More comprehensive Java-based analytical applications that can be run from a web page using Java Web Start, or downloaded to a user's workstation and run as a stand-alone application.

ANALYTICAL TOOLS

BLAST similarity searches (17) are available that provide standard web-based BLAST output as well as a web-based version of our BLAST parsing and visualization application, XS-BLAST (XML-SQL BLAST). XS-BLAST utilizes the XML-output option of the NCBI BLAST executable, and parses the results storing them in our SQL Server database. The result set is then provided to the user as an HTML table, as well as from a Java-based graphical visualization tool (Figure 3). Both BLAST searches are run locally on a Sun Solaris server and provide several customized databases for searching. These include all complete genomic *Poxviridae* nucleic acid sequences, all predicted *Poxviridae* protein sequences and all available *Poxviridae* nucleotide sequences extracted from GenBank.

Comparative analysis of the coding potential of all *Poxviridae* genomes is available as both pairwise and multi-genome comparisons. Both tabular listings of orthologous protein sets are provided as well as a graphical gene synteny plot of shared orthologs between any two genomes (Figure 3). Ortholog determination is based on an all versus all BLAST search of all poxvirus proteins in the PBRC database.

AVAILABILITY

The PBRC is developed and maintained at the Department of Microbiology, University of Alabama at Birmingham and the Department of Biochemistry and Microbiology, University of Victoria. The PBRC is available at http://www.poxvirus.org.

The PBRC framework has also been used to support development of similar bioinformatics resources for other virus families. The Virus Bioinformatics Resource at http://www.virology.ca contains databases and analytical tools to support research on coronaviruses, herpesviruses and baculoviruses.

FUTURE PLANS

In the near future, the PBRC will become part of a more comprehensive Viral Bioinformatics Resource Center for Biodefense (VBRC) that will include additional viruses listed as priority pathogens by the NIAID. In addition to expansion of the database and available analytical tools, the VBRC will make available more comprehensive help screens and tutorials that will provide users with step-by-step guides in how to use the site to answer typical questions that might be of interest to the laboratory researcher. The VBRC is being established as one of the NIAID's 'Bioinformatics Resource Centers for Biodefense and Emerging or Re-Emerging Infectious Diseases'. For more details on these centers, see http://www.niaid.nih.gov/dmid/genomes/brc/default.htm. The VBRC is available at http://www.biovirus.org.

ACKNOWLEDGEMENTS

This work was funded through two grants from NIAID/DARPA: U01 AI48706 (E.J.L.) and U01 AI48653-02 (R.M.L.B. and C.U.) and Canadian NSERC grant OPG0155125-01 to C.U.

REFERENCES

1. Henderson,D.A., Inglesby,T.V., Bartlett,J.G., Ascher,M.S., Eitzen,E., Jahrling,P.B., Hauer,J., Layton,M., McDade,J., Osterholm,M.T. *et al.* (1999) Smallpox as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *J. Am. Med. Assoc.*, **281**, 2127–2137.
2. Guarner,J., Johnson,B.J., Paddock,C.D., Shieh,W.J., Goldsmith,C.S., Reynolds,M.G., Damon,I.K., Regnery,R.L. and Zaki,S.R. (2004) Monkeypox transmission and pathogenesis in prairie dogs. *Emerg. Infect. Dis.*, **10**, 426–431.
3. Enserink,M. (2003) Infectious diseases. U.S. monkeypox outbreak traced to Wisconsin pet dealer. *Science*, **300**, 1639.
4. Upton,C., Slack,S., Hunter,A.W., Ehlers,A. and Roper,R.L. (2003) Poxvirus Orthologous Clusters (POCs): toward defining the minimum essential poxvirus genome. *J. Virol.*, **77**, 7590–7600.
5. Ehlers,A., Osborne,J., Slack,S., Roper,R.L. and Upton,C. (2002) Poxvirus Orthologous Clusters (POCs). *Bioinformatics*, **18**, 1544–1545.
6. Upton,C., Hogg,D., Perrin,D., Boone,M. and Harris,N.L. (2000) Viral genome organizer: a system for analyzing complete viral genomes. *Virus Res.*, **70**, 55–64.
7. Hiscock,D. and Upton,C. (2000) Viral Genome Database: storing and analyzing genes and proteins from complete viral genomes. *Bioinformatics*, **16**, 484–485.
8. Brodie,R., Roper,R.L. and Upton,C. (2004) JDotter: a Java interface to multiple dotplots generated by Dotter. *Bioinformatics*, **20**, 279–281.
9. Brodie,R., Smith,A.J., Roper,R.L., Tcherepanov,V. and Upton,C. (2004) Base-By-Base: single nucleotide-level analysis of whole viral genome alignments. *BMC Bioinformatics*, **5**, 96.
10. Hoskins,J., Alborn,W.E.,Jr, Arnold,J., Blaszczak,L.C., Burgett,S., DeHoff,B.S., Estrem,S.T., Fritz,L., Fu,D.J., Fuller,W. *et al.* (2001) Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.*, **183**, 5709–5717.
11. Glass,J.I., Lefkowitz,E.J., Glass,J.S., Heiner,C.R., Chen,E.Y. and Cassell,G.H. (2000) The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature*, **407**, 757–762.
12. Chen,N., Danila,M.I., Feng,Z., Buller,R.M.L., Wang,C., Han,X., Lefkowitz,E. and Upton,C. (2003) The genomic sequence of *Ectromelia* virus, the causative agent of mousepox. *Virology*, **317**, 165–186.
13. Wheeler,D.L., Church,D.M., Federhen,S., Lash,A.E., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E., Tatusova,T.A. *et al.* (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.*, **31**, 28–33.
14. Ball,L.A. and Mayo,M.A. (2004) Virology Division News: report from the 33rd Meeting of the ICTV Executive Committee. *Arch. Virol.*, **149**, 1259–1263.
15. Fauquet,C.M. and Mayo,M.A. (2001) The 7th ICTV report. *Arch. Virol.*, **146**, 189–194.
16. Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, D258–D261.
17. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.