

AD _____

Award Number: DAMD17-02-2-0059

TITLE: Development of a Viral Biological-Threat Bioinformatics Resource

PRINCIPAL INVESTIGATOR: Elliot J. Lefkowitz, Ph.D.

CONTRACTING ORGANIZATION: The University of Alabama at Birmingham
Birmingham, Alabama 35294-0111

REPORT DATE: October 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20040105 185

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE October 2003	3. REPORT TYPE AND DATES COVERED Annual (30 Sep 2002 - 29 Sep 2003)	
4. TITLE AND SUBTITLE Development of a Viral Biological-Threat Bioinformatics Resource			5. FUNDING NUMBERS DAMD17-02-2-0059	
6. AUTHOR(S) Elliot J. Lefkowitz, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Alabama at Birmingham Birmingham, Alabama 35294-0111 <i>E-Mail:</i> elliotl@uab.edu			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES Original contains color plates: All DTIC reproductions will be in black and white.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 Words) In response to the potential use of viruses as biological weapons, we have established the Viral Biological-threat Bioinformatics Resource (VBBR) that collects, catalogs, annotates, and analyzes genetic information related to potential viral threats. This work expands upon available knowledge of virus replication, pathogenicity, and virus-host interactions on the basis of individual protein domains, individual genes, and whole genomes. To date, we have constructed a genome and gene sequence database that has been populated with the sequence information for viruses currently listed on the NIH and CDC priority pathogen list. We have also developed a variety of analytical and visualization tools that aid in the analysis of the genomic information coded for by these viruses. Finally, the information developed as a result of this work has been made available to the scientific community through a (currently access-controlled) web site (http://vbbr.genome.uab.edu) that supports research efforts to develop environmental detectors, diagnostics, antiviral compounds, new vaccines, and animal models in support of biodefense research goals.				
14. SUBJECT TERMS Biodefense, biological terrorism, bioinformatics, viruses, sequence database			15. NUMBER OF PAGES 13	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

Table of Contents

Cover.....	1
SF 298.....	2
Table of Contents.....	3
Introduction.....	4
Body.....	4
Key Research Accomplishments.....	13
Reportable Outcomes.....	13
Conclusions.....	13
References.....	
Appendices.....	

INTRODUCTION:

In response to the potential use of viruses as biological weapons, we have established the Viral Biological-threat Bioinformatics Resource (VBBR) that collects, catalogs, annotates, and analyzes genetic information related to potential viral threats. This work expands upon available knowledge of virus replication, pathogenicity, and virus-host interactions on the basis of individual protein domains, individual genes, and whole genomes. This information has been made available to the scientific community to support current research efforts to develop environmental detectors, diagnostics, antiviral compounds, new vaccines, and animal models.

Specifically we are:

1. developing a relational database that supports the data storage, annotation, analysis, and information exchange goals of this proposal;
2. developing a world-wide web site that will allow all information and analytical results compiled and generated by this work to be publicized to the scientific community, providing rapid access to specific information as required by an individual researcher;
3. collecting existing gene and genomic sequences for viral threat agents and importing them into the database for subsequent annotation and analysis;
4. providing computer-automated and human-directed annotation to confirm and update existing biological information in the sequence records for all sequences; and
5. performing a variety of analyses on genome and gene sequences to provide additional information on their structure, function, and evolution.

BODY:

Substantial progress has been made in accomplishing many of the tasks outlined in the original statement of work. Accomplishments extend to database construction, data population, development and implementation of analytical and visualization tools, and publication of all available information on a web site. The accomplishments are listed below, itemized according to the original statement of work and task list.

Task 1. To develop a relational database that will support the data storage, annotation, analysis, and information exchange goals of this proposal. (Months 1-6)

- The database has been created based on our previous work on poxviruses and has been updated and refined to better support a wider range of virus threats. An overview of the database schema is provided in figure 1.

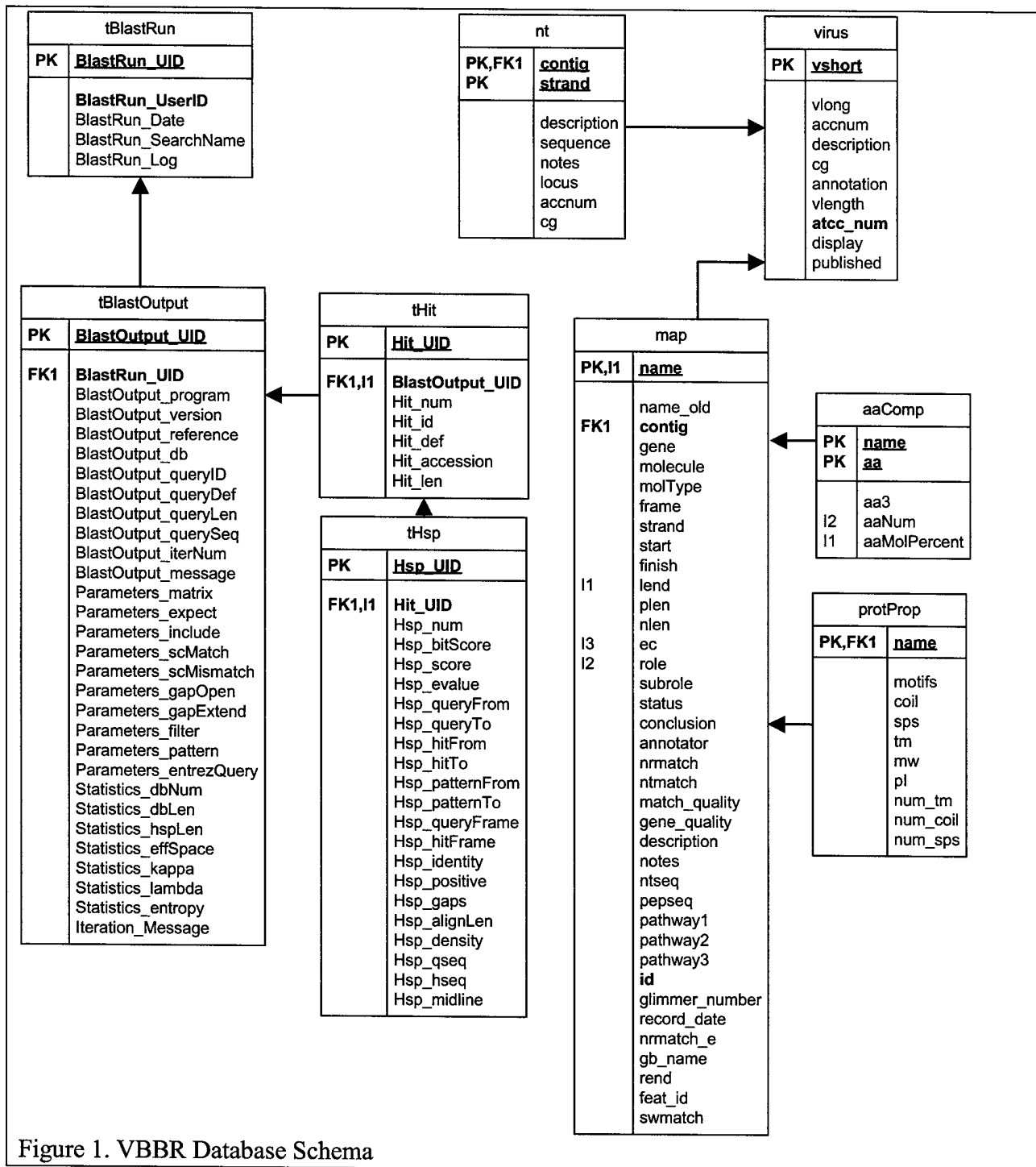


Figure 1. VBRR Database Schema

Task 2. To develop a world-wide web site that will allow all information and analytical results compiled and generated by this work to be publicized to the scientific community, providing rapid access to specific information as required by an individual researcher. This web site will also provide the necessary security and access controls to ensure that only individuals and groups as designated by the granting agency (USAMRMC) can utilize these resources. (Months 1-12)

- The web site has been established and is accessible at the url: <http://vbbr.genome.uab.edu>.
- Login to the web site requires registration. Following registration, a user account is established, and access is secured via a username/password. To register, a user needs to

logon to the following web page: <https://btd.genome.uab.edu/register/register.asp> using the following credentials: Username: genuab; Password: \$mgbf2003 and fill out the form. The user will be vetted for access and when approved will then be notified via Email when the account is established.

- The home page for the VBBR web site is shown below in figure 2.

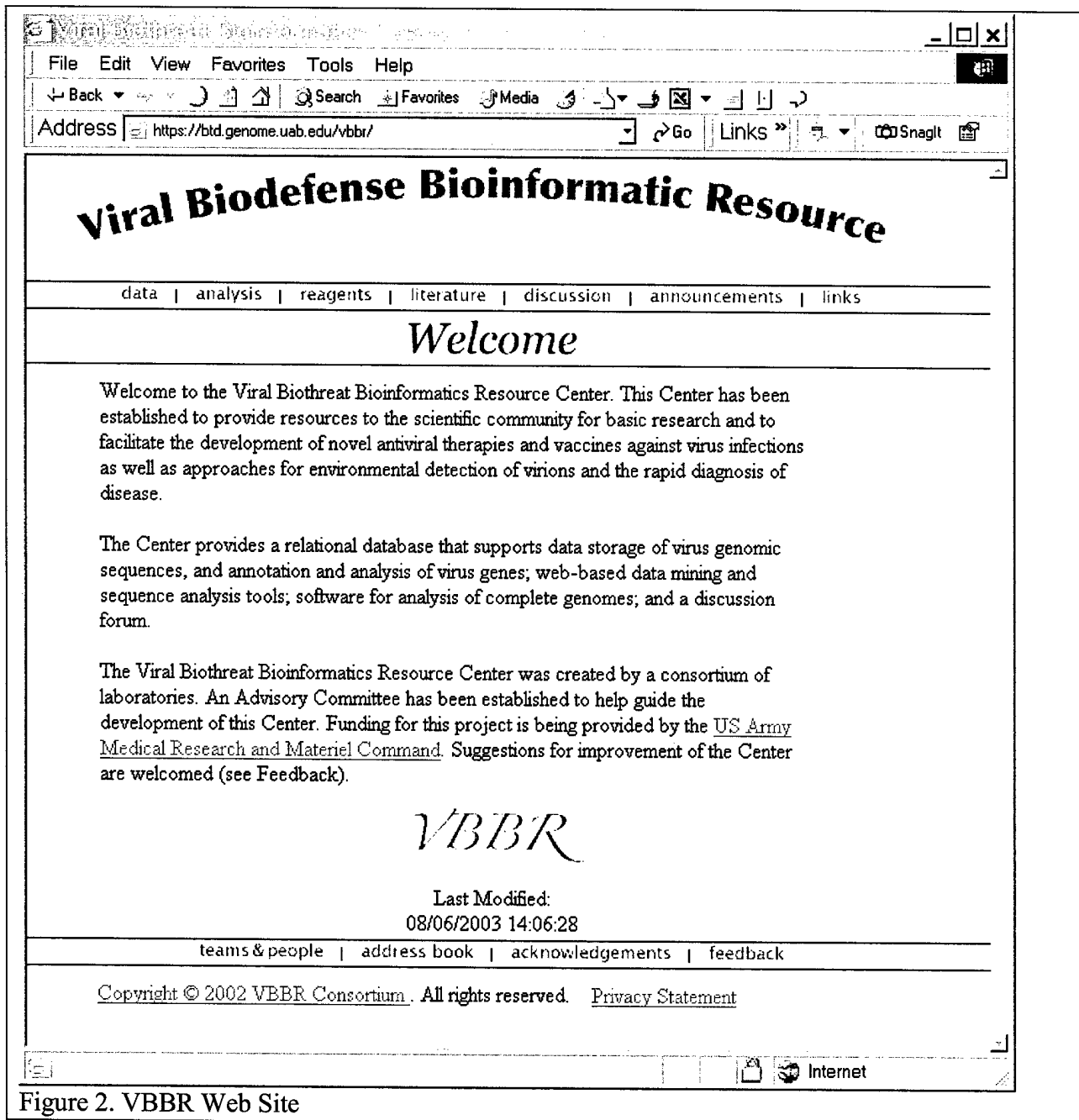


Figure 2. VBBR Web Site

Task 3. To collect existing gene and genomic sequences for viral threat agents and import them into the database for subsequent annotation and analysis (Months 1-24).

- a. We will obtain existing gene and genomic sequences for all viruses that potentially could be used as agents of biological warfare or terror. The**

information collected will include the corresponding descriptive annotation provided with the sequences.

- An automated parser has been developed (XSeq Exchange) that can parse XML-formatted GenBank sequence records and their annotation and load that information into the VBBR database. This application will greatly enhance our ability to add new viral genomes to the VBBR database. The database schema for the parser is shown below in figure 3.

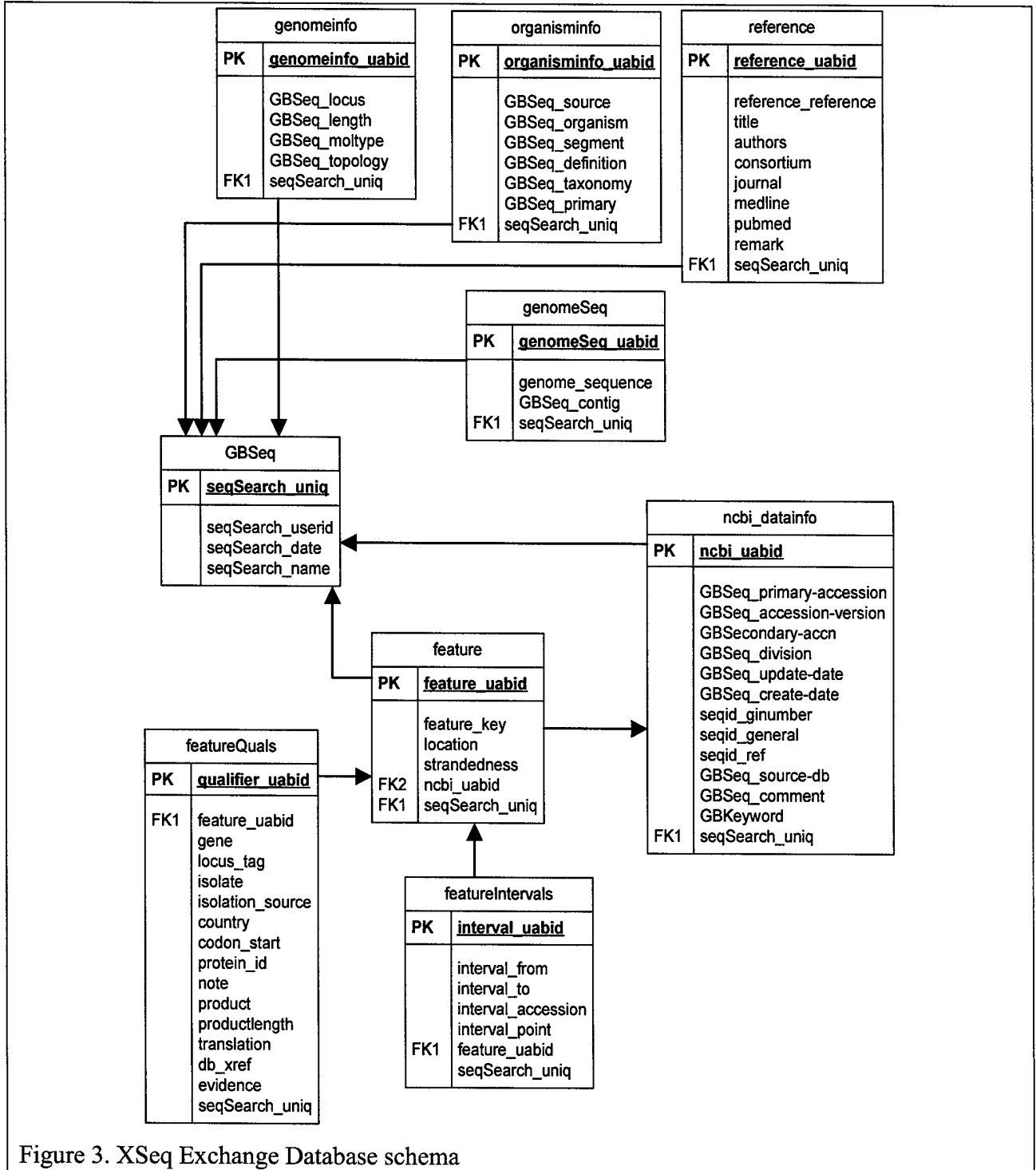


Figure 3. XSeq Exchange Database schema

- The genome and gene database is currently populated with most of the viruses currently on the NIAID A-C priority pathogen list. These are shown on the genomes page of the VBBR web site as displayed in figure 4.

VBBR

data | analysis | reagents | literature | discussion | announcements | links | home

Data

[Genomes](#) | [Genes](#) | [Sequence Download](#)

Complete Genomes Available in the Viral Biodefense Bioinformatics Resource

[Click for information on changes to the VBBR virus designations.](#)

Virus (Click for Gene Map)	Complete Genomes	Genome Size (bases)	Accession Number (GenBank NCBI)	Accession Number (EMBL EBI)
Andes virus segment L, complete sequence	Y	6,562	NC_003488	
Andes virus segment M, complete sequence	Y	3,671	NC_003467	
Andes virus segment S, complete sequence	Y	1,871	NC_003466	
Crimean-Congo hemorrhagic fever virus S gene for nucleoprotein, genomic RNA	N	1,672	AJ010649	
Crimean-Congo hemorrhagic fever virus gene for envelope glycoprotein precursor, complete cds, isolate:88166	N	5,367	AB069675	
Eastern equine encephalitis virus, complete genome	Y	11,675	NC_003899	
Guanarito virus strain INH-95551 segment S, complete sequence	Y	3,343	AY129247	
Human hepatitis virus type A RNA, complete genome	Y	7,474	M20273	
Hantaan virus S segment gene for nucleocapsid protein, complete cds, strain:Chen4	Y	1,623	AB027101	
Hantaan virus gene for putative polymerase, genomic RNA	Y	6,533	X55901	
Hantaan virus M segment gene for polyprotein, complete cds, strain:NC167	Y	3,626	AB027115	
Junin virus GPC and N genes for glycoproteins and nucleocapsid protein, complete cds	N	3,400	D10072	
La Crosse virus segment L, complete sequence	N	6,960	NC_004108	
La Crosse virus segment M, complete sequence	N	4,527	NC_004109	
La Crosse virus segment S, complete sequence	N	984	NC_004110	
Lassa virus segment L, complete sequence	Y	7,279	NC_004297	
Lassa virus segment S, complete sequence	Y	3,402	NC_004296	
Lymphocytic choriomeningitis virus segment L, complete sequence	Y	6,660	NC_004291	
Lymphocytic choriomeningitis virus segment S, complete sequence	Y	3,376	NC_004294	
Machupo virus strain Carvallo segment S, complete sequence	Y	3,439	AY129248	
Marburg virus, complete genome	Y	19,112	NC_001608	
Tick-borne encephalitis virus, complete genome	Y	11,141	NC_001672	
Puumala virus genomic S-RNA, isolate: Tobetsu-60Cr-93, complete sequence, viral-complementary strand	Y	1,833	AB010731	
Rabies virus, complete genome	Y	11,932	NC_001542	
Reston Ebola virus, complete genome	Y	18,891	NC_004161	
Rift Valley fever virus S segment, complete sequence	Y	1,690	NC_002045	
Rift Valley fever virus L segment, complete sequence	Y	6,608	NC_002043	
Rift Valley fever virus, complete genome	Y	3,885	NC_002044	
SARS coronavirus, complete genome	Y	29,751	NC_004718	
Venezuelan equine encephalitis virus, complete genome	Y	11,444	NC_001449	
Western equine encephalomyelitis virus, complete genome	Y	11,484	NC_003908	
West Nile virus, complete genome	Y	10,962	NC_001563	
Yellow fever virus, complete genome	Y	10,862	NC_002031	
Zaire Ebola virus, complete genome	Y	18,959	NC_002549	

teams & people | address book | acknowledgements | feedback

Copyright © 2002 VBBR Consortium. All rights reserved. [Privacy Statement](#)

Figure 4. VBBR Genomes Page.

- b. We will also collect and provide the references for publications in the scientific literature that describe studies on the structure and function of orthopoxvirus genes and genomes. We will link this information on our web site to the appropriate gene records in our database.**
 - Not yet done. Tasked for year 2.
- c. We will initially obtain viral sequence from public databases such as GenBank. We will also obtain unpublished sequence information through inquires and interactions with other scientists that have been initiated by our collaborators.**
 - Not yet done. Tasked for year 2.

***Task 4.* To provide computer-automated and human-directed annotation to confirm and update existing biological information in the sequence records for existing as well as newly obtained sequences. (Months 6-24)**

- Computer-automated annotation has been completed for the viruses listed in figure 4. Human directed annotation is tasked for year 2. An example annotation record is shown in figure 5.

***Task 5.* To perform a variety of analyses on genome and gene sequences to provide additional information on their structure, function, and evolution. (Months 6-24)**

- Initial computer-aided analysis has been completed. Additional analyses tasked for year 2.

***Task 6.* To provide through the web site, a set of analytical and visualization tools that will allow users of this resource to mine the data for useful information, to perform comparative analyses between these and other viruses, and to better visualize and assess the significance of their results. (Months 6-24)**

- Currently available analytical and visualization tools include:
 - a. Genome map visualization (Figure 6.)
 - b. Genome protein gene ortholog comparisons
 - c. Gene synteny visualization (Figure 7.)
 - d. BLAST similarity searches
 - e. BLAST search parsing and database storage
 - f. Java-based Visualization of BLAST search results (Figure 8.)
- A new database and web server has been installed and is now supporting all VBBR resources. This new server provides added security and has greatly increased performance of the VBBR web site.
- Development of additional analytical and visualization tools are tasked for year 2. These will include tools for creation and visualization of multiple sequence comparisons and tools for the inference and visualization of evolutionary trees.

Data

[Genomes](#) | [Genes](#) | [Sequence Download](#)

VBBR Gene Record for: REBOV-PENN:RNA-dependent RNA polymer

Last updated by dba On 7:30:2003 3:23:53 PM



Virus: Reston Ebola virus strain Pennsylvania
 Genbank Locus Name: RNA-dependent RNA polymer
 Gene Name: RNA-dependent RNA polymer
 VBBR Gene Number: REBOV-PENN_16
 Molecule: Protein
 Strand: Forward
 Protein Length: 2212
 Nucleotide Length:
 Start: 11550
 Finish: 18188
 Molecular Weight: 252546.62
 pI: 8.15
 Description: RNA-dependent RNA polymerase [Reston Ebola virus]

Protein Statistics	
Amino Acids:	Composition
Number of putative Transmembrane Domains:	0 (Click for the Complete TransMem Record)
Prosite Motifs:	No Hits
Coiled-coil Segments:	None
Secretory Signal Peptides:	None

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

Notes

Protein Sequence

```

MATOHTQYED ARLSSEPIVLD QCDLVTRACG LYSSYSINPQ LRQCKLPKHI YRLKFDITVS KFLSDTPVAT LPIDYLVPII IRLSTGHGDR PLTPTCNOFL DELINVTIHD
AAFLDYVYKA TGAQDHLTNI ATREKLNKEI LNNDYVHOLF FVHDLILAR RGRNLRGNRR STWVFDHFI DILGYGDYIF WKIPLSLIPV TIDGVPHAAT DUYQTLFKE
SILGHSQILS VSTAELLIMC KDIIITCRFNT SLIASIAKLE DVDVSDYDDP SDILKIYNAG DVVISILGSE GYKIIKYLEP ICLAKIQOLCS KFTERRGREL KFDGHSVIND
LRELISNRRL KDYOQEKIRD FHKILLQLOL SPOQFCELES VQKHWGHPIL HSEKATQKVK RHATILKALR PNVIETFCV FKYNIKHYF DSQGTWYSVI SDRNLTPGLN
SFIKRNHFFS LPMIKDLLVE FYHLNHPPFL STKVISDLSI FIKDRATAVE QTCWDAVFEP NVLGYNPPNK FSTKRVPPOF LEQEDFSES VLNVAQELHY LLPQNRNFSF
SLKEKELNIG RTPGKLPYLT RNVQTLCEAL LADGLAKAFP SNMNVVIERE QKESILHQAS WHHTSDDFGE NATVRGSSVF TDLEKYNLAF RYEFYAPFIE YCNHCYGVNR
VFVNHVYLLP QCYHNSDYF NEFHNVNLGN REYFPEGPS YRGLGGIEG LQOKLWTSIS CAQISLVEIK TGFKLRSAVM GDNQCITVLS VFPLKTDPEE QEOGAEDNAA
RVAASLAKVT SAGCFLKED ETVHSGEYV FQKQVNLGV OLPQSLKTA A RMAPLSDAIF DDLQGLASI GAFERAISE TRHILPCRIV AAFHTYFAVR ILQYHHLGFN
KGIIDLGOISL SKFLDYGITL LTLAVPOVIG CLSFLNPKC FFRNFGDPT SCLFQLRVYL EMVNHKDLFC PITSKNPNC SAIDFVLNFS GLNYPGSQDL TSEFLRQVRR
SITLTARNKL IMTLFHASAD LEDEMCKWL ISSNPFVMSRF AADIFSRTPS GKRLQILGYL ECTRLIASK IINWNSETF LDKLRKTLQ RNLWFSVLD HCDLLADAL
QKISCTVDLA QILREYVWSH ILEGRLIGA TLPCHVEQFK VKVLGQVEPC PECLNKKGSN AYVSAVVDQ VVSAWPTSR ISWTICSGVP YIGSRTEKQ GQPAIKRCP
S3ALKEAIEL ASRLTVVTQG GSNSEQLIRP FLEARVNLSV SEVLQMTPSH YSGNIVHRVN DOYSPHSFMA NRMNNTATRL IVSNTLGEF SGGQAARDS NIIFQNVINL
AVALYDIRFR NTNTSDIRHN RAHLHLTECC TKEVPAQYLT YTSALNLDLS RYRDNELIYD SNPLKGLNCL NLTIDSPVVK GPRLNMIEDD LIRFPHISGW ELAKTVQSI
ISDINSSTID PISSEGETRSF TTFELTYPOI GLLYSFGAVL CFYLGNTILW TKKLDVEQFL YYLHNQHLNL PHERALRVFKP TFKHASVMSR LHEIDSNFSI YIGGTSGDRG
LSDAARFLR TALASLOFL KSWIIDROKT IPLWIVYPLE GQQPESINEF LHKILGLLQK GPKSIPKEVS IQNDGHDLA ENNVYVNSKS TASNFFHSL AYVRSRKRK
TDDHDFSRG DSTLTDFVRK FSSNHQSDER YYNVTCGRSP KPOERKDFSQ YRLSNNQTM SHRRKKGKFF KWNPCKMLM SORGTVITEG DYFQNTPTP DDVSSPHRLI
LFFKLGHNH HAHDOAQEL MNQNIQVYH OLRSMIDTTI YCRFTGVSS MHYKLDVLL EYNSFDSAIT LAEGEGSGAL LLLQKYSTRL LFLNLTATEH SIESEVWSGF
STPRMLPTM QVHEQGVV ILNNSAQIT DITSEHLSW QVWLPQVE IYMDAETTE NLNRSQLYRA VNLILDHID FOYLKVVVK VFLSDIEGLI WINDYLAPLF
SGAVLIKPII SSARSSEVYL CLSNLISTNR RSAROTHKAC LGVIRDALQA QVQRGVLS HLAQYATKAL HCEYIGLGF SLEKLYVHRV NLVDTGLGFL SSVIRHLTL
QAEIRDLVD YHLMRESRTO TYHFIKTAGR RITKLVNDFL KFLSIQALK NNSSVYTELK KLPEVINCN RPYHTNCEC QEKFFVQTLY LQRLDAEIK LIERLTGLHR
HT
    
```

Nucleotide Sequence

[teams & people](#) | [address book](#) | [acknowledgements](#) | [feedback](#)

Figure 5. VBBR Gene Annotation Record.

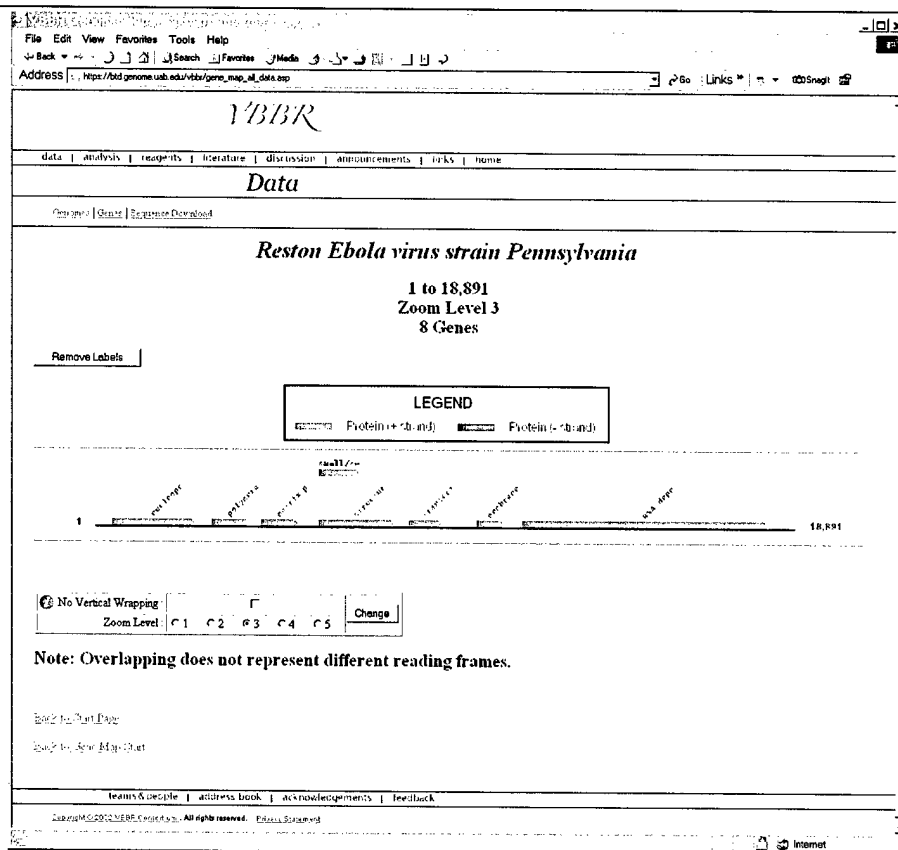


Figure 6. VBBR Genome Map.

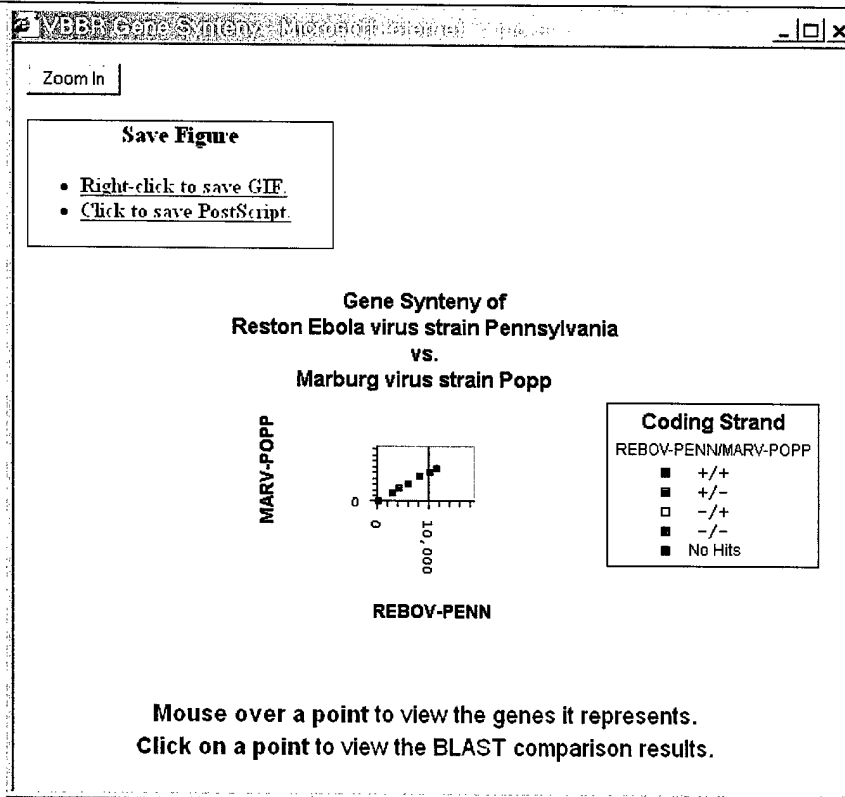


Figure 7. VBBR Gene Synteny Map.

BLAST Viewer
Exit About

PROGRAM: blastx
DATABASE: nr
QUERY SEQUENCE: sars

Sort By: Bit Score
e-Value Filter: e-100 to 0 e-10 to e-100 .01 to e-10 Select All

Hit Selection: Horizontal Zoom: Level 1 Level 2 Level 3 Level 4
 Hide Checked Hits Show All Hits

Number of Nucleotides: 1 4000 8000 12000 16000 20000 24000 28000 28715

gi 26007546 zef		ORF1ab polyprotein [Murine hepatitis virus]
gi 26008095 zef		transmembrane domain 2 [Bovine coronavirus]
gi 22219303 pfb		Chain A, Structure of Coronavirus Main Proteinase
gi 25121564 zef		coronavirus nsp3 (ND2) [Murine hepatitis virus]
gi 26008087 zef		coronavirus nsp5 [Bovine coronavirus]
gi 25121551 zef		coronavirus nsp6 [Avian infectious bronchitis virus]
gi 25121568 zef		coronavirus nsp7 (6EL); growth-factor-like protein
gi 25121569 zef		coronavirus nsp9 (RdRp); RNA-dependent RNA polyme
gi 6625761 gb A		RNA-directed RNA polymerase [murine hepatitis viru
gi 26008091 zef		coronavirus nsp10 (NB, NTPase/HEL); metal-binding
gi 25121571 zef		coronavirus nsp11 [Murine hepatitis virus]
gi 26008094 zef		coronavirus nsp13 [Bovine coronavirus]
gi 5565844 gb A		spike glycoprotein [murine hepatitis virus]
gi 6706916 gb A		spike glycoprotein [bovine coronavirus]
gi 6179905 gb A		membrane protein [murine hepatitis virus]
gi 11640712 gb		nucleocapsid protein [Equine coronavirus]

BLAST Viewer Usage:
CLICK on HIT NAME to view sequence record.
CLICK on HIT GRAPHIC to view alignment(s).

Java Applet Window

Figure 8. VBBR BLAST Viewer. The figure shows the results of a BLASTX search of the SARS genome against the Genbank non-redundant protein database.

KEY RESEARCH ACCOMPLISHMENTS:

- Establishment of the VBBR Database
- Establishment of the VBBR Web Site
- Development of the XSeq Exchange GenBank sequence parser
- Population of the VBBR database with virus genomic information
- Computational annotation of virus genes
- Implementation of analytical and visualization tools for BLAST searches
- Implementation of analytical and visualization tools for genome ortholog comparisons

REPORTABLE OUTCOMES:

Informatics:

- VBBR Database
- VBBR Web site (<http://vbbr.genome.uab.edu>)
- VBBR analytical and visualization tools

Presentations:

- USAMRIID/WRAIR/LANL/UAB Workshop, August 11, 2003 WRAIR
- USAMRMC Bioinformatics Workshop, November 4, 2003 (Poster)

Funding proposals submitted extending this work:

- NIH/NIAID Proposal in response to the RFP: NIH-NIAID-DMID-04-34, "Bioinformatics Resource Centers for Biodefense and Emerging/Re-emerging Infectious Diseases"

CONCLUSIONS:

Development of the Viral Biological-threat Bioinformatics Resource (VBBR) has substantially followed the original task list and essentially all tasks proposed for year 1 have been accomplished. The result is a Bioinformatics Resource Center that can now begin to support the needs of basic and applied biodefense research directed at gaining a better understanding of the pathogenesis, evolution, and overall biology of viral pathogens as well as support the development of detectors, diagnostics, antivirals, and vaccines. Future work will follow the original task list and is aimed at providing human-curated gene records along with additional analytical and visualization tools that can better support understanding the role of individual genes in virus pathogenicity.

REFERENCES:

None other than the VBBR web site.

APPENDICES:

None