

COOPERATIVE RELATIONAL DATABASE INITIATIVE FOR THREAT REDUCTION

Michelle Sheahan and Luther E. Lindler
Department of Bacterial Diseases
WRAIR
Silver Spring, MD 20910

ABSTRACT

In order to create a resource for basic and clinical research in biological threat reduction, we have developed an annotated relational database. This database is comprised of gene sequences from public databases and researchers' laboratory results, the Bioterrorism Defense Database runs on a Microsoft SQL platform and is accessible on a password-protected Internet site. The Biodefense database project is a collaborative effort between the Walter Reed Army Institute of Research, the US Army Medical Research Institute of Infectious Diseases, the Los Alamos National Laboratory, and the University of Alabama at Birmingham.

INTRODUCTION

We have developed a relational database of genes relevant to the studies aimed towards biological threat reduction. The database is comprised of individual gene sequences with their amino acid translations that have been annotated with information about toxicity, available probes, antibiotic resistance, source organism, strain, and literature references. Gene sequences of toxins, virulence factors and antibiotic resistance are taken both from GenBank searches and from researchers' own unpublished sequence data. The database is accessible on a password-protected web site, and is searchable by various criteria including organism, gene name and accession number. The immediate goal of the Bioterrorism Defense Database creation effort has been to present microbial pathogen data to researchers in a format that is useful, clear and comprehensive. The unique feature of this database is the one gene-one sequence design and the way in which the information is compiled and annotated. Over the next year, we plan to include more extensive annotations for each gene sequence, prepared by expert curators. Our database is one facet of a large-scale biological threat portal that is a collaborative effort between researchers at USAMRIID (Kevin Anderson), WRAIR, DOE-CBNP (Gerald Myers and Electra Sutton) and the University of Alabama at Birmingham (Elliot Lefkowitz). Our vision is that the final portal include database information that is crucial to researchers performing studies in the area of biodefense.

METHODS

The Bioterrorism Defense Database is a gene-based relational database. Many of the entries are selected from publicly released gene sequences submitted to GenBank while others are unpublished laboratory sequences. The gene entries are identified by gene name, GenBank accession number, and the unique DNA sequence of the gene's specific coding region. Additional information on antibiotic resistance, toxins, virulence factors, and probes is added into each record, as well as links to references and the protein translation (Figure 1.) This enables the

Report Documentation Page

*Form Approved
OMB No. 0704-0188*

Public reporting burden for the 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 the 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. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 00 JAN 2002	2. REPORT TYPE N/A	3. DATES COVERED -			
4. TITLE AND SUBTITLE Cooperative Relational Database Initiative For Threat Reduction		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Bacterial Diseases WRAIR Silver Spring, MD 20910		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES This article is from ADA409494 Proceedings of the 2001 ECBC Scientific Conference on Chemical and Biological Defense Research, 6-8 March , Marriott's Hunt Valley Inn, Hunt Valley, MD., The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 4	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

researcher to search for homologies based on gene sequence, and to view an annotated record that can be further analyzed with compatible DNA or protein sequence analysis software.

Bioterrorism Defense Database

[Update/Delete This Record](#) | [Download Fasta Formatted Sequence](#)

Id:	12163
Accession Number:	AF217740
Name:	gyrA
Genus:	Yersinia
Species:	pestis
Strain:	632
Common Name:	plague
Virulence Factor:	No
Probes:	No
Toxin Producing?:	No
Antibiotic Resistance?:	Yes Display the Antibiotic Resistance Gene Record(s) (for this accession number)
Sequence Complete?:	No
Total bases in sequence:	492
Total bases in gene:	492
Start base:	1
Stop base:	492
Annotator:	GENDB/MSheahan
Record Date:	7/4/2000 11:52:42 AM
Reference:	Lindler,L.E. and Jahan,N. Nucleic acid detection of ciprofloxacin resistance using non-radioactive hybridization probes Unpublished; submitted 3/2000.
Comments:	(ciprofloxacin resistance) from clone 1-106; ciprofloxacin resistant mutant; sequence given is partial; product is gyrase A; db_xref GI 7264712; compared to sequence deposited in GenBank Accession Number AF217736
DNA Sequence:	atgtccgta tttcggagc tgcgttacc gatgtccgt atggactgaa accggtgcac cgtcccgtae tgtttggcat gaattgact ggtaagact ggaataaacc atacaaaaa tcggcccggt tagtcgggga cgttatcggt aaataccacc cgcattgtga cagccggtc taccacacta tcgtgggtat gcccagccg tctcactgc gctatatgt ggtggatggg cagggtaact tcggtccgt cgatgggtac tcgcccggy cgatgggtta taccgaaac cgtatgcta aaattgctca cgaattgta cgggatttag aaaaagatac cgttgacttc gtgctaact atgatggtac ggaacaatt ccggctgta tgcgaccag aatccctaac ctgctggtaa accgttcgtc gggtattgcy gtagggatgg caaccaatat tcgccacat aatctttctg ag
Protein Sequence:	MSVIUGRALFDVVDGLKPVHREVLFAENVLGNUNKPYRKSARV VGDVIGRYHPHCDSAVYDTIVRMAQFFSLRYMLVDGQGNFGSVDGDSAAAHRYTEIRM SKIAHELLADLEKDTVDFVPHYDGTQIPAVNPTRIPLNLLVNGSSGIAVGHATNIPPH NLSE

We have included genes and organisms that are relevant to the study of biological defense, including bacterial and viral threat agents. The organisms and diseases used in our current search criteria are shown in Table 1.

TABLE 1. Organisms/diseases/genes of interest.

Anthraxis	Junin
Anthrax	Lassa
apamin	Machupo
Batrachotoxin	Marburg
Beta-bungarotoxin	Notexin
Botulism	pilin
Brucella	Ricin
Chloramphenicol resistance	Rift Valley
Clostridium perfringens toxin	Sabia
Conotoxin	Salmonella toxin
Coxiella	Salmonella virulence
Crimean-Congo	Salmonella pathogenicity
curare	Saxitoxin
Dengue	SEB
Diamphotoxin	Shiga toxin
Diphtheria toxin	Shigella
Ebola	Strep Resistance
EEEV	T2 toxin
Escherichia coli toxin	Taipoxin
Escherichia coli pathogenicity	Tet Resistance
Escherichia coli virulence	Tetanus Toxin
fimbrillin	Tetrodotoxin
Francisella	Topoisomerase
Guanarito	Vaccinia
Hantavirus II	Variola
Hantavirus I	VEENCGR
Heat-Labile	Vibrio cholerae
Heat-stabile	WEEV
	Yersinia pestis
	Yersinia enterocolitica

Data is entered into the Bioterrorism Defense Database either manually, by cutting and pasting from researchers' gene sequence results, or automatically, by reading web-accessible data with a parse application. This parse application was developed in collaboration with the Los Alamos National Laboratory and makes downloading web-accessible gene databanks more efficient and accurate. Once the fields of the gene entry page are populated, the annotator reviews the entries, makes any changes or corrections, and adds information from further analyses. Antibiotic resistance, toxin, and probe data is entered as separate tables within the database. Our plans include organizing the gene entries into clusters and adding the functionality of protein and gene analysis tools linked directly to the gene entry pages. The annotator will thus be able to complete additional analyses and predictions for each set of clustered gene products, enhancing the information available to the community of researchers using the Bioterrorism Defense Database.

Users can search the database with a BLAST routine using protein or nucleotide sequences, and view annotated search results. All-against-all search capability will be added in the near future. In addition, the Bioterrorism Defense Database is part of a larger effort to organize and present relevant data to the research community through a portal website. This Tri Agency Chemical and Biological National Security Program Portal is the result of a bioinformatics collaboration between Lawrence Livermore National Laboratory, Los Alamos National Laboratory, and USAMRIID/WRAIR. The CBNP portal will include the capability of simultaneously searching several participating biological threat databases, including the Bioterrorism Defense Database presented here, so that the researcher has access to the most comprehensive, up-to-date annotated analyses from experts in the field of biological terrorism defense.

CONCLUSIONS

Over the past year, we have worked to streamline the data entry process so that the most basic information about relevant genes can be accurately and efficiently added to the database. The next step is making the Bioterrorism Defense Database more valuable and informative to the researcher, by both expanding the breadth of information on our sequences of interest, and on improving its graphical presentation. With our collaborators at USAMRIID, DOE-CBNP, and the University of Alabama at Birmingham, we envision that the database will be a valuable source of data on the following:

- genes, transcripts, and gene products
- genomes and plasmids
- homologies: orthologies, paralogies, xenologies
- regulatory elements and repeats
- pathogenicity islands
- primers and probes
- molecular signatures; fingerprints
- recombinant constructs
- mechanisms of pathogenicity
- antibiotics and resistance
- growth properties; phenotypic data
- variability; alignments and cluster analyses
- protein structures; immunological properties
- geographical distribution and backgrounds
- clinical and host data
- prophylaxis and treatment
- literature