

Immune Epitope Database NEWSLETTER

Volume 5, Issue 2

<http://www.immuneEPITOPE.org>

April 2008

Facilitating Direct Data Submission

One of the IEDB's goals is making the process of submitting data more user-friendly, since many labs find the current XML format to be daunting without the help of IT-savvy personnel. The IEDB staff is implementing a two-step approach to achieve this end. In late 2008 we will be testing and releasing template files that are simplified versions of the IEDB data structure. The concept is that the user will choose a template file that is designed to hold experimental data of a particular "type", and download it to their system. Using that template, which has a spreadsheet format, users copy/paste their data into the template, and then upload that back into the IEDB for semi-automated validation prior to public release as a searchable record(s).

For release in 2009, the IEDB staff is working on an interactive system that will guide users as they enter data into the IEDB. This system is being designed to combine the best aspects of simple spreadsheet-style entry with logic-driven error checking. In addition, users will be able to combine bulk entries and smaller experimental data sets. If you are interested in beta testing this system, please contact the IEDB at contact@immuneEPITOPE.org.

Inside this Issue

- 1.) Facilitating Direct Data Submission
Implementing A User-Friendly Process
- 2.) Curation Update
Up Next in the Queue
- 3.) Curator Milestone
Efforts of the Curation Staff
- 4.) Looking Back to 2007
- 5.) Looking Forward to 2008
Launch of IEDB 2.0
- 6.) Recent Publications
Authored by Team members
- 7.) Epitope Cluster Analysis Tool
- 8.) Upcoming Events
IEDB Exhibit Booths

Looking Back to 2007

Curation Update

Curation of data relating to NIAID Category A, B, and C priority pathogens, NIAID emerging and re-emerging infectious diseases (<http://www3.niaid.nih.gov/research/topics/emerging/list.htm>), Malaria, Hepatitis B, Clostridium tetani, Leishmania, Candida albicans, and herpesvirus is current for articles appearing in PubMed as of the end of December 2007. These will be updated in April 2008 to cover newer published articles and pertinent references recently brought to our attention. The IEDB curation team has started on allergen references and other infectious diseases. All reference categories will continue to be updated quarterly. Curation of autoimmune diseases will start later in 2008 upon completion of the allergy category. Users are invited to bring references to our attention that are potentially relevant to the IEDB but do not appear in the database. References that are deemed to meet the IEDB criteria for curation will be queued for processing in accordance to our NIH-directed priorities.

Curator Milestone

April 2008 marks the three year anniversary of the IEDB's first full-time curators. On April 6, 2005, three biochemists started their careers at LIAI in a new scientific venture that would eventually grow to include twelve full-time curators and a document specialist. Two of the three original hires are still at LIAI. Nima Salimi now serves as the IEDB Curation/Database Manager and Romulo DeCastro, Ph.D. has the role of Senior Curator. Thanks to their early efforts and the efforts of the rest of the curation staff, well over 4000 literature references have now had their epitope data captured in the IEDB.

During calendar year 2007, the quantity of data available in the IEDB increased significantly with the addition of over 2000 fully curated references and over 30,000 records from direct data submissions from Large Scale Epitope Discovery contractors. The IEDB also gained traction as a resource for researchers, with 27 articles citing it. In addition, a number of new features were introduced to the website during 2007:

- Tree structure for allele browser
- Browse by 3D structure
- Patent page
- Epitope dataset forum
- New XML format for export
- Antibody prediction tools BepiPred and DiscoTope
- Artificial Neural Networks (ANN) methods for T cell epitope prediction expanded to allow sequence lengths greater than nine
- MHC Class II epitope prediction methods expanded from only ARB to ARB, Smm_align, Sturniolo, and consensus methods
- Epitope cluster analysis Tool

Looking Forward to 2008 - IEDB 2.0

A “new and improved” Immune Epitope Database website will be making its appearance this June. Users will see a website that is significantly more user-friendly. Usability experts from Virginia Tech were enlisted in May 2007 to assess the IEDB website from a human-machine interface perspective. Most of their 90 suggestions have been incorporated in

continued on page 5

Recent Publications

Analysis of epitope information related to *Bacillus anthracis* and *Clostridium botulinum*.

Zarebski LM, Vaughan K, Sidney J, Peters B, Grey H, Janda KD, Casadevall A, Sette A.
Expert Rev Vaccines. 2008 Feb;7(1):55-74.

PMID: 18251694

We have reviewed the information about epitopes of immunological interest from *Clostridium botulinum* and *Bacillus anthracis*, by mining the Immune Epitope Database and Analysis Resource. For both pathogens, the vast majority of epitopes reported to date are derived from a single protein: the protective antigen of *B. anthracis* and the neurotoxin type A of *C. botulinum*. A detailed analysis of the data was performed to characterize the function, localization and conservancy of epitopes identified as neutralizing and/or protective. In order to broaden the scope of this analysis, we have also included data describing immune responses against defined fragments (over 50 amino acids long) of the relevant antigens. The scarce information on T-cell determinants and on epitopes from other antigens besides the toxins, highlights a gap in our knowledge and identifies areas for future research. Despite this, several distinct structures at the epitope and fragment level are described herein, which could be potential additions to future vaccines or targets of novel immunotherapeutics and diagnostic reagents.

Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines

Bui HH, Sidney J, Li W, Fusseder N, Sette A.
BMC Bioinformatics. 2007 Sep 26;8:361.

PMID: 17897458

In an epitope-based vaccine setting, the use of conserved epitopes would be expected to provide broader protection across multiple strains, or even species, than epitopes derived from highly variable genome regions. Conversely, in a diagnostic and disease monitoring setting, epitopes that are specific to a given pathogen strain, for example, can be used to monitor responses to that particular infectious strain. In both cases, concrete information pertaining to the degree of conservancy of the epitope(s) considered is crucial. **RESULTS:** To assist in the selection of epitopes with the desired degree of conservation, we have developed a new tool to determine the variability of epitopes within a given set of protein sequences. The tool was implemented as a component of the Immune Epitope Database and Analysis Resources (IEDB), and is directly accessible at <http://tools.immuneepitope.org/tools/conservancy>. **CONCLUSION:** An epitope conservancy analysis tool was developed to analyze the variability or conservation of epitopes. The tool is user friendly, and is expected to aid in the design of epitope-based vaccines and diagnostics.

Epitope Cluster Analysis Tool

In this issue of the IEDB Newsletter, we highlight one of the new tools introduced in 2007, the Epitope Cluster Analysis tool. This tool groups epitopes into clusters based on sequence identity.

A cluster is defined as a group of sequences which have a sequence similarity greater than the minimum sequence identity threshold specified.

To use the tools, users enter epitope sequences directly in a text area or they upload them from a specified file. If uploading data from a file, the file contents will appear in the text area. Two acceptable sequence formats are PLAIN and FASTA. A sequence in PLAIN format is separated by a new line. A sequence in FASTA format begins with a single-line description, followed by line(s) of sequence data. The description line is distinguished from the sequence data by a greater-than (" $>$ ") symbol in the first column. The user next selects the sequence identity threshold at which they want to calculate epitope clusters. Clicking the "Submit" button starts the calculation. The "Reset" button clears all input parameters. Example input is shown in Figure 1. On output, clusters are displayed in a table format where clusters are indicated by table rows which have the same color. All calculated cluster results can be saved to a file by clicking on the "Download data to file" button (Figure 2).

To use this tool, please visit the IEDB Analysis Resource page:

<http://tools.immuneepitope.org/main/jsp/menu.jsp>

Figure 1 - Example of Input

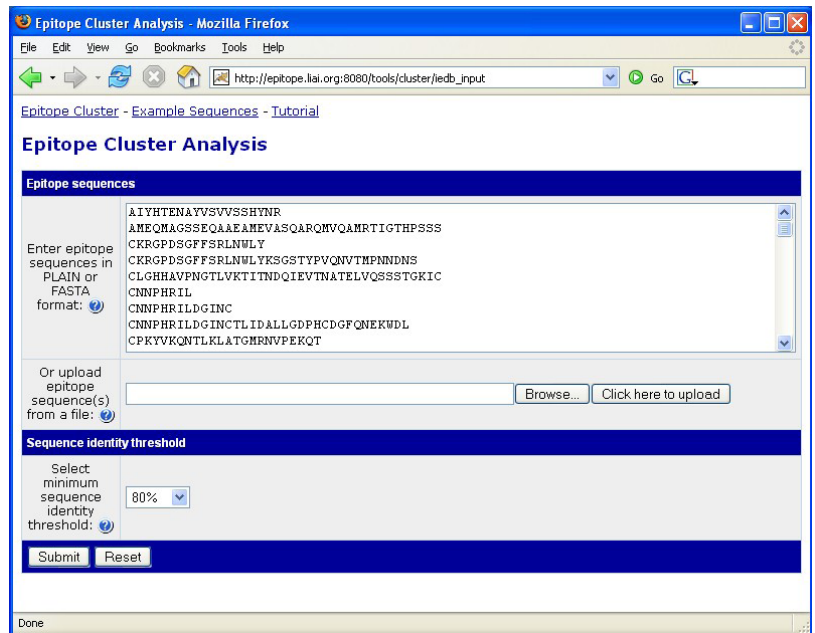
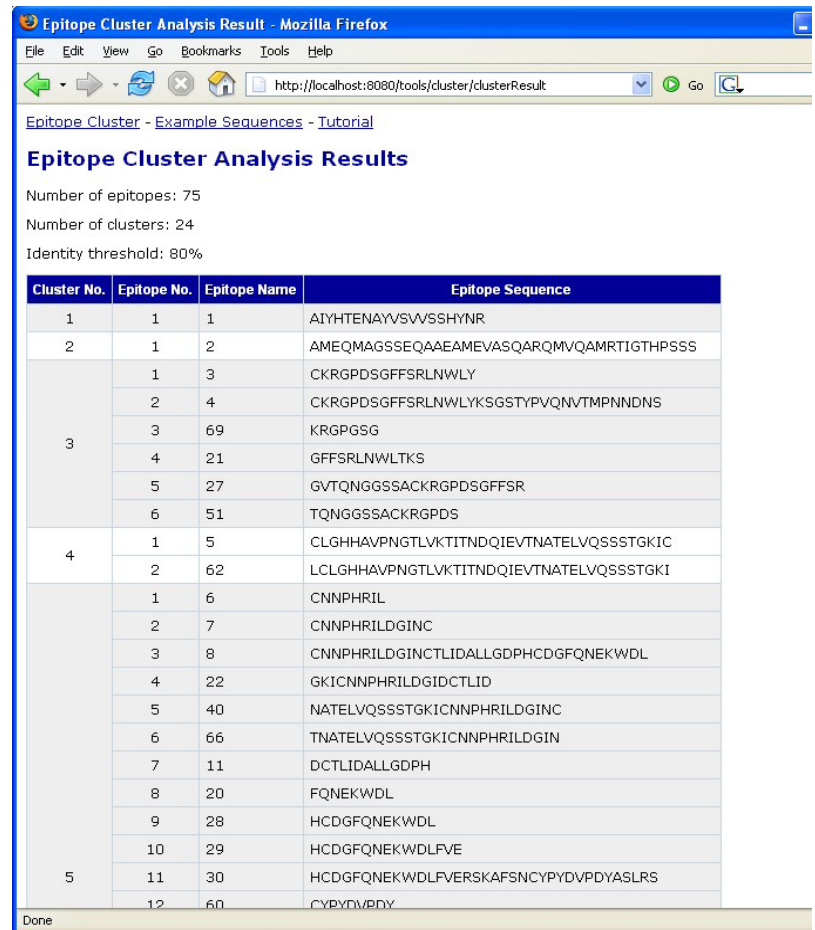


Figure 2 - Example of Output



Upcoming Events

Exhibit Booths

Vaccine Research

11th Annual Conference

May 5-7, 2008

Baltimore Marriott Waterfront Hotel
Baltimore, MD

Booth #10

American Society for Microbiology

108th General Meeting

June 1-5, 2008

Boston Convention & Exhibition Center
Boston, Massachusetts

Booth # 548

Stop by to visit us!

Contact Information

The Immune Epitope Database & Analysis Resource is supported by a contract from the National Institute of Allergy & Infectious Disease, NIH, DHHS (Contract HHSN266200400006C). The newsletter is distributed four times a year. We welcome communication from the users of the IEDB database and invite suggestions for articles in future issues. To subscribe to the IEDB newsletter or to contact project staff, send your email information to the email address below.

Immune Epitope Database and Analysis Resource
c/o La Jolla Institute for Allergy & Immunology
9420 Athena Circle
La Jolla, CA 92037
(858) 752-6500

Email: contact@immuneEPITOPE.org
Web: <http://www.immuneEPITOPE.org>

Principal Investigator:

Alessandro Sette, Ph.D. - alex@liai.org

Co-Principal Investigator:

Bjoern Peters, Ph.D. - bpeters@liai.org

Project Director:

Stephen Wilson, Ph.D. - swilson@liai.org

Production:

Jody Dang
Ward Fleri, Ph.D.

continued from page 2

the site redesign. Under the hood, there will be a new data schema that better describes the data and better accommodates data curated from allergy and autoimmune references. We will also be able to accurately represent passive immunity for the first time. During the process of mapping data from the old schema to the new, data fields have been checked for consistency, evidence codes have been added, and the overall quality of the data has been improved.