

NCBI News, August 2008

Peter Cooper, Ph.D.¹ and Dawn Lipshultz, M.S.²

Created: August 26, 2008; Updated: November 25, 2008.

Featured Resource: New Graphical Sequence Viewer

NCBI recently released a new sequence viewer that allows smooth zooming and browsing of records in a horizontal graphical format. Multiple horizontal tracks can display overlapping features such as assembly details, genes, transcripts, coding regions, protein products and polymorphisms. A multi-panel display also can show views of varying scales from hundreds of megabases at the chromosome level to a single gene view, or all the way down to the individual bases of the sequence. It also produces a text sequence display suitable for copying and pasting sequence to other applications. The sequence viewer is available as the “Graphics” display for nucleotide and protein sequences, the “Sequence Viewer” link from Gene and polymorphism (SNP) records, and from the Genes map in the NCBI Map Viewer.

The default display in the sequence viewer consists of two panels. The top panel provides an overview of the entire sequence. This overview panel has a search feature that allows you to find features on the sequence and navigate to them. The wrench icon present on all graphical panels provides access to the configuration controls to change the color scheme, graphical format, and add content layers. The second default panel shows a graphical view (Figure 1) of the region highlighted in the overview. You can add additional graphical views by clicking the icon with the colored bars in the uppermost left of the overview. On each graphical panel a set of icons at the top provides navigational controls including the ability to scroll left or right along the sequence (arrows) and to zoom in or out (plus and minus magnifying glasses.) The “ATG” button in the center is a short-cut that zooms directly to the individual bases of the sequence. You can generate a text view of the region displayed in a graphical panel by clicking the “ATGC” icon in the upper left-hand-corner. The text view is color coded to match the features displayed in the corresponding graphics (genes, coding regions, exons, introns.)

The viewer is particularly well suited to displaying large and complex sequences with many features. The figure shows a region of human chromosome 6 displayed in two panels in the sequence viewer. This region contains three members of the tumor necrosis factor family, lymphotoxin alpha (LTA), tumor necrosis factor (TNF) and lymphotoxin

¹ NCBI; Email: cooper@ncbi.nlm.nih.gov. ² NCBI; Email: lipshult@ncbi.nlm.nih.gov.

beta (LTB). The lower panel has been configured to display polymorphisms and shows the nucleic acid, corresponding protein sequence and codons. The sequence viewer also features a simplified URL structure that makes it convenient to link to specific views. For example, to see human chromosome 6 in the viewer simply load the following URL in a Web browser:

www.ncbi.nlm.nih.gov/projects/sviewer/?id=NC_000006.10

You can quickly find the tumor necrosis factor gene family members by searching for “lymphotoxin alpha” in the search box in the top panel and following the link to LTA.

The NCBI sequence viewer is powerful new platform for displaying complex sequence features in an easily navigable system. This viewer will continue to evolve and is destined to assume a more central role as a general sequence display and navigation tool on the NCBI Web site.

New Databases and Tools

Primer-BLAST

The new Primer-BLAST tool helps users design more effective and gene-specific PCR primers. Primer-BLAST will design primers for an input template target and will also check the specificity of input primers. The tool uses code from the highly regarded Primer3 to help design primers, then uses a specialized BLAST search to eliminate primers that match non-target sequences. To get started with Primer-BLAST go to: www.ncbi.nlm.nih.gov/tools/primer-blast/.

Bookshelf

The Bookshelf has added three new books entitled: *PubMed Clinical Q&A*, *Drug Class Reviews*, and *Advances in Patient Safety: From Research to Implementation*. Books can be found at: www.ncbi.nlm.nih.gov/sites/entrez?db=Books.

Microbial Genomes

39 finished microbial genomes were released (June 12-August 13). The original sequence data files submitted to GenBank/EMBL/DDBJ can be found at: <ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/>. The RefSeq provisional versions of these genomes are available via FTP at: <ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>.

GenBank News

GenBank release 166.0 is available via web and FTP that includes information as of June 11, 2008. GenBank has reached a milestone of 200 billion basepairs.

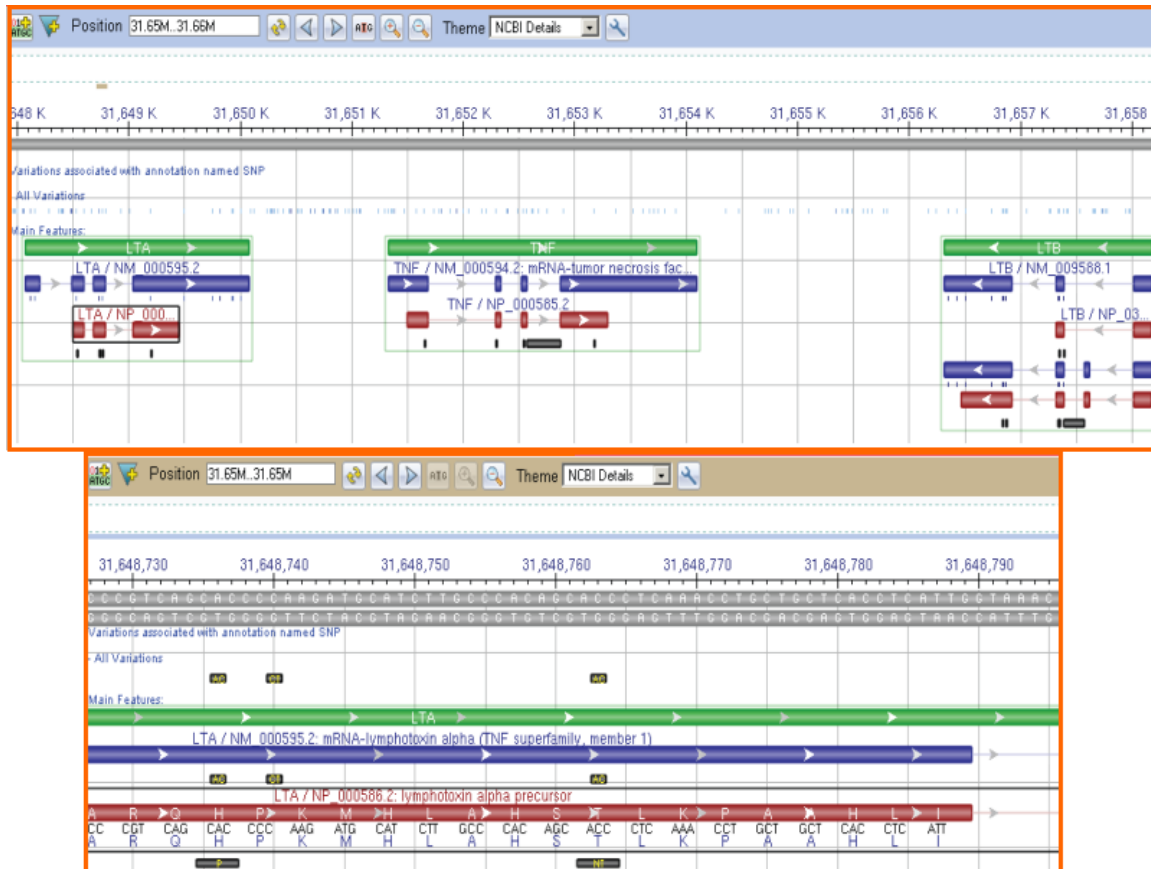


Figure 1. Two detailed graphical panels from the new NCBI sequence viewer showing the tumor necrosis factor gene family members on human chromosome 6 (NC_000006.10). The lower panel is the zoomed-in display on the second exon of the lymphotoxin alpha gene showing nucleic acid sequence, protein sequence, codons, and polymorphisms.

Updates and Enhancements

CDTree

CDTree, NCBI's versatile protein domain hierarchy viewer, has released version 3.1. The new version is now supported on Apple computers running OSX and higher. Other new features include an annotation matrix viewer, multi-CD operations, sequence tree coloring for greater flexibility, and higher efficiency in working with large hierarchies. Find more information at: www.ncbi.nlm.nih.gov/Structure/cdtree/cdtree.shtml.

Clone Finder

The Clone Finder tool has been updated and supports mouse build 37, with more organisms coming shortly. Supported organisms will have a clone finder icon on the Map Viewer home page. Improvements to Clone Finder include: the addition of new libraries; improved searching and download mechanisms; and additional annotation. The mouse

Clone Finder page is located at: www.ncbi.nlm.nih.gov/projects/mapview/mvhome/mvclone.cgi?taxid=10090.

RefSeq

RefSeq Release 30 is available via web and FTP. This full release incorporates genomic, transcript, and protein data available as of July 7, 2008 and includes 8,572,852 records, 5,590,368 proteins, and sequences from 5,395 organisms. The RefSeq homepage is: www.ncbi.nlm.nih.gov/RefSeq/.

Genome Assembly

An updated genomic assembly of the red flour beetle, version 2.1, was released in June. Zebrafish Zv7, pig Sscrofa5, and horse EquCab2.0, were released in July. Cow build 4.0 and pea aphid build 1.0 were released in August. For more annotation updates and links to Genome Resource pages see: www.ncbi.nlm.nih.gov/Genomes/.

Announce Lists and RSS Feeds

Fourteen topic-specific mailing lists are described on the Announcement List summary page. Announce lists provide email announcements about changes and updates to NCBI resources. http://www.ncbi.nlm.nih.gov/Sitemap/Summary/email_lists.html

Seven RSS feeds are now available from NCBI including news on PubMed, PubMed Central, NCBI Bookshelf, LinkOut, HomoloGene, UniGene, and NCBI Announce. <http://www.ncbi.nlm.nih.gov/feed/>

Comments and questions about NCBI resources may be sent to NCBI at: info@ncbi.nlm.nih.gov, or by calling 301-496-2475 between the hours of 8:30 a.m. and 5:30 p.m. EST, Monday through Friday.