U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Cooper P, Morris R. NCBI News, August 2011. 2011 Aug 31 [Updated 2011 Sep 1]. In: NCBI News [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 1991-2012.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/

NCBI News, August 2011

Peter Cooper, Ph. D.[1] and Rana Morris, Ph.D.[2]

Created: August 31, 2011; Updated: September 1, 2011.

NCBI Discovery Workshops September 27-28 at NLM: Seats still available

NCBI will present a two-day workshop September 27 and 28, 2011, on the NIH campus in Bethesda, Maryland. The course is free and is open to anyone interested in NCBI resources. The four workshops are Sequences, Genomes, and Maps; Proteins, Domains and Structures; NCBI BLAST Services; and Human Variation and Disease Genes. These workshops provide hands-on experience exploring practical examples using tools and databases on the NCBI website. The Discovery Workshops page has more details and a link to register for the course.

Feature Highlight Now Available in Sequence Databases

A new way to highlight annotated sequence features, named sites, and regions listed in the FEATURES table is now available in the Entrez sequence databases. The new tool is helpful in visualizing the extent and location for such important features as genes, coding regions, exons, and mRNAs in nucleotide sequences and conserved domains, modification sites, and interaction sites in protein sequences.

Activating the Feature Highlight

Figure 1 shows an example of the new feature highlight for an exon in the MAOA gene (NG_008957). Clicking on any of the hyperlinked items in the left-hand column of the FEATURES table section of a sequence record displayed in the Entrez Nucleotide or Protein services highlights the corresponding region on the sequence. Single segment features -- for example, an exon (Figure 1), or multiple segment features -- for example, mRNA alignments on genomic DNA (Figure 2), can be highlighted. The highlighted segments are displayed with white residue letters and a brown background. The Highlight Bar and also opens at the bottom of the page that provides additional information and controls.

The Feature Highlight Bar

The Feature Highlight Bar provides details about the highlighted region, controls for navigating additional features on the record, and has links to display the highlighted regions as separate sequences for downloading and further analysis. The Details box that is open by default on the Highlight Bar shows the detailed annotation from the FEATURES table for the now highlighted region. The Details box can be collapsed if desired by clicking on the Details link. Clicking the link again re-opens the box. The number of highlighted segments is shown at

Author Affiliations: 1 NCBI; Email: cooper@ncbi.nlm.nih.gov. 2 NCBI; Email: morrisrc@ncbi.nlm.nih.gov.

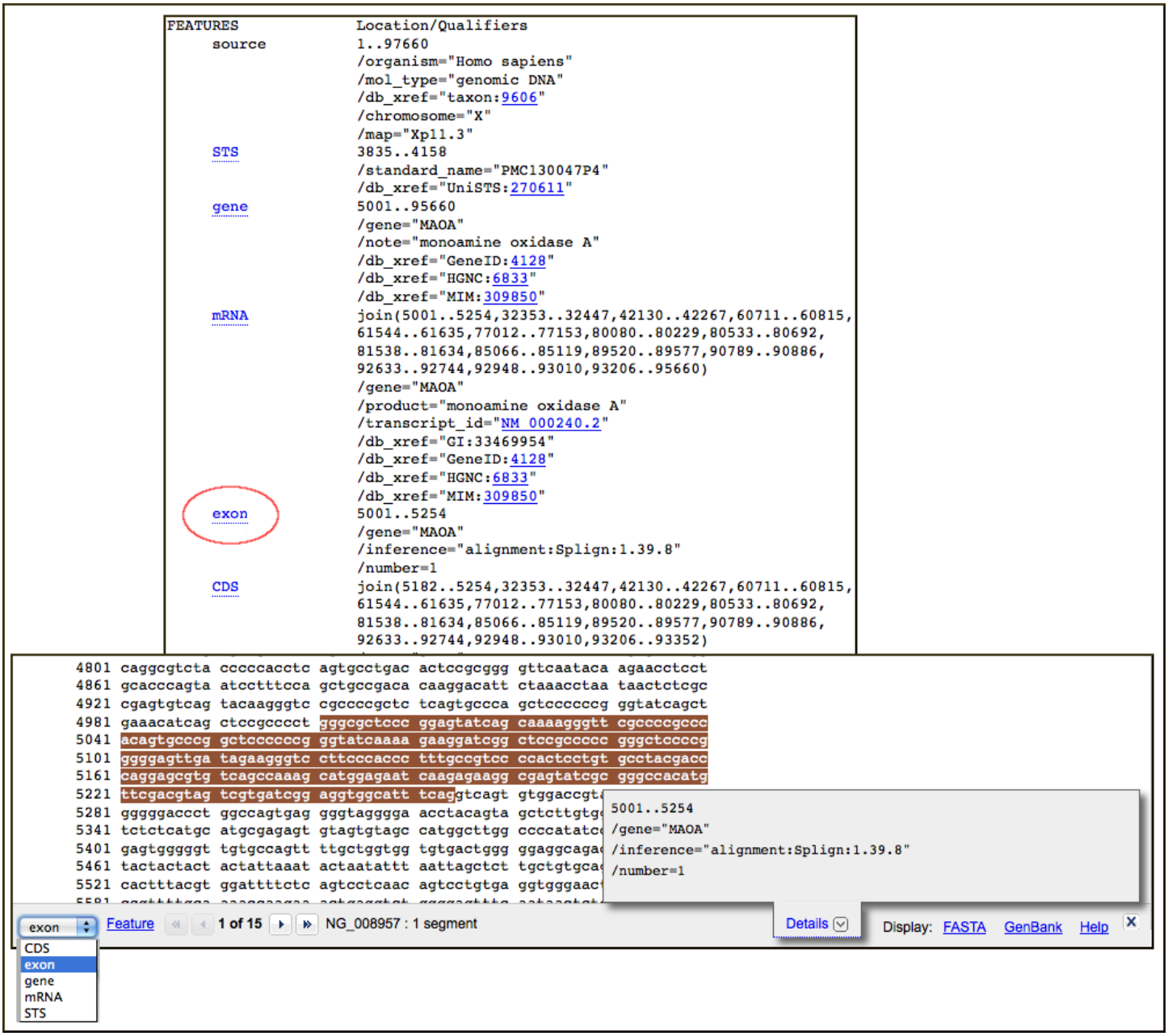

Figure 1. Feature highlighting shown for an exon feature on the NCBI RefSeq Gene record for Monoamine Oxygenase A (NG_008957). Clicking on the exon link in the left-hand column of the FEATURES table activates highlight and opens the Highlight Feature Bar at the bottom of the page. Other feature types can be highlighted by selecting them from the Feature pull-down list. Clicking the Feature link returns to the FEATURES table of the record. The number of features of the selected type is shown – 15 in the case of exon features for NG_008957. Clicking the navigational arrows allows jumping to the next or previous feature of a given type. The details box, which may be closed if desired, re-states the range and qualifiers for the highlighted feature. The FASTA and GenBank links display the highlighted region as a separate view available for copying, downloading, or submitting for further analysis.

the right of the sequence accession in the Highlight Bar. In the example in Figure 2, opposite strand features are indicated with the notation “minus strand” to the right of the number of segments on the Highlight Bar.

Navigating Using the Feature Highlight Bar

If there is more than one feature of the same type, the navigational arrows on the bar allow jumping to the next, previous, first, and last instances of that feature. The Feature pull-down list at the left-hand side of the bar allows

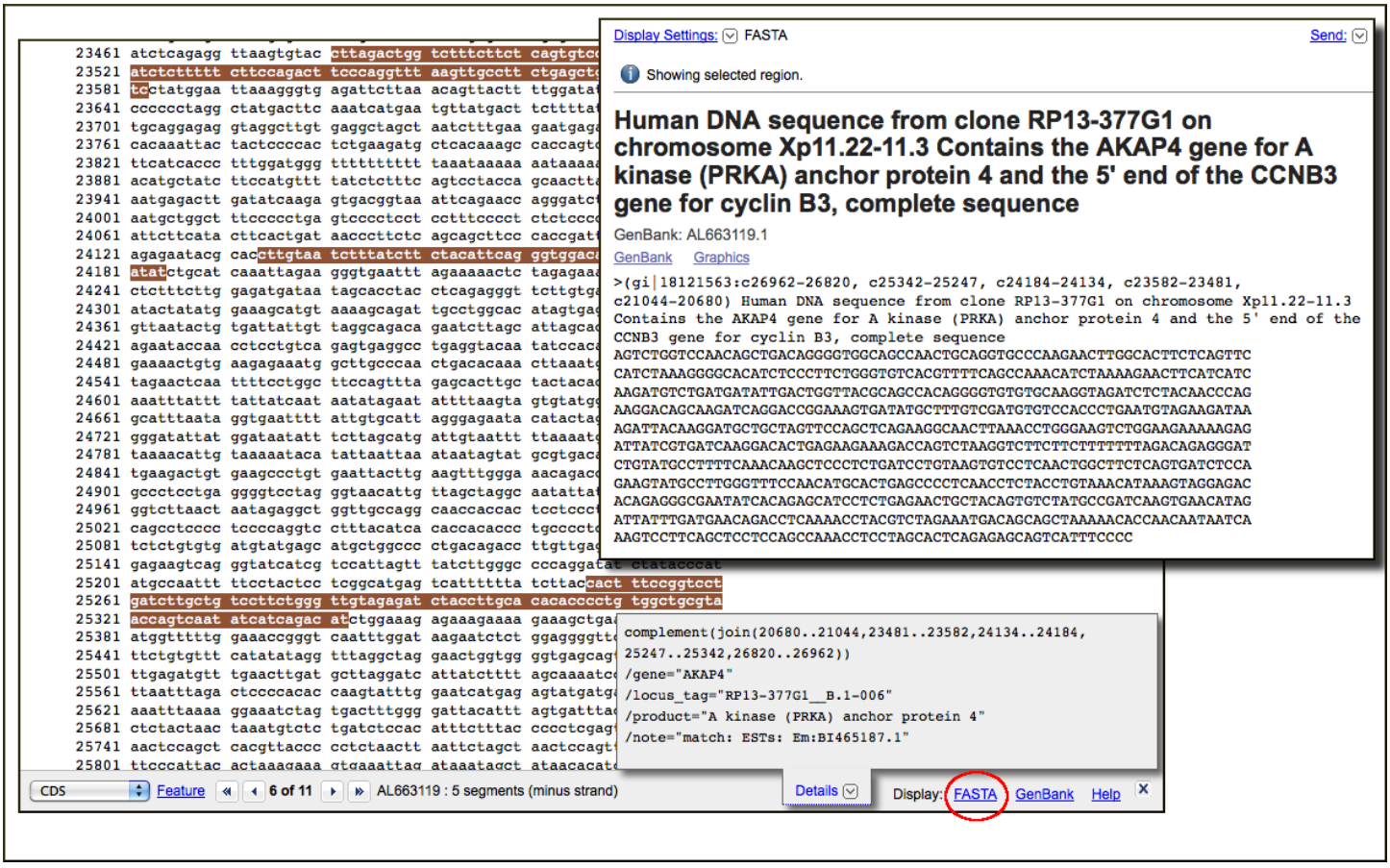

Figure 2. Highlighting and displaying a multi-segmented mRNA feature on the minus strand of a BAC clone sequence (AL663119). *Back panel.* Highlighting a splice variant of AKAP4 gene that is the sixth mRNA feature on the record. There are 5 highlighted segments on the minus strand of the record. Clicking the FASTA link displays the corresponding region shown in the *Front panel.* The complementary strand is shown automatically giving the same sequence as the mRNA.

selecting other available feature types. The highlight moves to the next available instance of the selected feature type. The Feature link returns the display to the corresponding position in the FEATURES table of the record.

Displaying Highlighted Regions as Separate Sequences

The FASTA and GenBank links on the right-hand side of the bar present the highlighted sub-sequence in the these formats in the Nucleotide or Protein Entrez system and provide a simple means to display and download the corresponding sequence or to forward it to the available analysis tools: BLAST, Primer-BLAST, Find in this Sequence, and Identify Conserved Domains (protein only). As shown in Figure 2, the sequence displayed in FASTA format is the appropriate strand for the feature, in this case the complementary or minus strand of the record.

Summary

The new ability to highlight features in sequence records complements the Find-in-Sequence tool described in the September 2010 NCBI News and adds powerful new visualization and search options to the NCBI sequence database.

New videos on NCBI's YouTube channel

Three new videos are available on NCBI's YouTube Channel. Two instructional videos show how to display the six-frame translations of a DNA sequence in the graphical sequence viewer on the web (Sequence Viewer: Six Frame Translations) and in the standalone Genome Workbench annotation and analysis tool (Genome Workbench: Six Frame Translations).

The PMC 10th Anniversary video, celebrating the ten years of the PubMed Central online public access full-text database, now joins the two other anniversary celebrations: the NCBI 20th anniversary video and the collection of talks from the GenBank 25th Anniversary.

Updated Genome Workbench (v2.4.0)

An update to NCBI's Genome Workbench (v2.4.0) is now available. Genome Workbench is a standalone sequence viewer, annotation and analysis platform. The new version has many new features, improvements, and a few bug fixes that are described in the release notes. The latest Genome Workbench pre-compile packages and source code are available from the download page.

Conserved Domain Database updated (v2.31)

Version 2.31 of the Conserved Domain Database (CDD) is now available. The new release contains 292 new or updated NCBI-curated domain models and now includes domains from SMART version 6. The CDD data are searchable in the Entrez and BLAST services at the NCBI Website and are available for download from the FTP site.

Microbial Genomes Update

One hundred fifty-five finished microbial (archaeal and bacterial) genomes were released during June, July and August 2011. The original sequence data files submitted to International Sequence Database Collaboration (INSDC) are available in the Bacteria directory in the genomes area of the GenBank FTP site. RefSeq provisional versions were released for a selected set of 86 of the complete INSDC microbial genomes during the same period. These are available from the /genomes/Bacteria directory on the FTP site.

In addition, 317 microbial whole genome-shotgun (WGS) sequencing projects were added to the INSDC during this period. The original submitted files are available in the Bacteria_DRAFT directory in the GenBank genomes area. RefSeq provisional versions of 86 WGS microbial projects were released in the /genomes/Bacteria_DRAFT area of the FTP site.

All GenBank and RefSeq microbial genomes are incorporated in the NCBI integrated Entrez search and retrieval system and the BLAST sequence similarity search service.

GenBank News

GenBank release 185 is available through the NCBI web and FTP sites. The current release incorporates data available as of August 14, 2011 and, with the whole-genome shotgun portion, contains 338,987,064,933 bases from 207,281,745 sequence records. Release notes describe the current state of data and upcoming changes.

RefSeq News

RefSeq Release 48 is available through Entrez, BLAST, and the RefSeq FTP site. The current release includes 18.2 million Reference Sequence records from 12,235 different organisms. The RefSeq release notes provide more detailed information.

NCBI will no longer archive new sequencing data from The Cancer Genome Atlas (TCGA)

NCBI will no longer archive new sequencing data from The Cancer Genome Atlas (TCGA) in the Sequence Read Archive. The release of dbGaP's TCGA study phs000178 version 5 on August 15, 2011 constitutes the up-to-date and final compendium of files available at NCBI.

The Growth of PubChem

PubChem now contains over 30,000,000 chemically unique compounds with over 500,000 bioassays. Research labs, institutes, organizations, and companies have submitted over 85,000,000 substances over the past seven years. Steady growth in content and usage is expected to continue. The PubChem news page has more details.

New Simple Object Access Protocol (SOAP)-based BLAST service

A new Simple Object Access Protocol (SOAP)-based service is available. The SOAP interface can be used to develop applications that interact with the NCBI BLAST web service to submit searches and retrieve results. Documentation and links to the Web Service Definition Language (WSDL) and sample clients are available on the NCBI Bookshelf.

NCBI at the ICHG/ASHG Meeting in Montreal: Workshop on Medical Genetics

NCBI scientists will present a special workshop at the combined International Congress of Human Genetics and American Society for Human Genetics meeting at Montreal Convention Center on October 12th at 12:30 P.M. The workshop entitled, "Genetics and Medicine at the National Center for Biotechnology Information (NCBI)", will provide information on genome-scale resources for medical genetic genetics available at the NCBI including finding and downloading data, analysis, and management of data sets. NCBI will also staff an exhibit booth at the meeting.

Announce Lists and RSS Feeds

Seventeen topic-specific mailing lists are available that provide email announcements about changes and updates to NCBI resources including dbGaP, BLAST, GenBank, and Sequin. The various lists are described on the Announcement List summary page. Subscribe to the NCBI Announce list to receive updates on the NCBI News.

Twenty-one RSS feeds are now available from NCBI including news on PubMed, PubMed Central, NCBI Bookshelf, LinkOut, HomoloGene, UniGene, and NCBI Announce.

NCBI's Facebook page and Twitter feed also provide updates on NCBI resources.

Send comments and questions about NCBI resources to info@ncbi.nlm.nih.gov, or call 301-496-2475 between the hours of 8:30 a.m. and 5:30 p.m. EST, Monday through Friday.