

NCBI News, June 2015

Tree Viewer version 1.5 improves performance

Tuesday, June 30, 2015

NCBI Tree Viewer version 1.5 includes several new features, improvements and bug fixes, including improved tree loading notifications, fixed resizing in full view and more. For a full list of updates, see the [Tree Viewer release notes](#).

NCBI Tree Viewer is a tool for viewing your phylogenetic tree data.

June 3rd webinar "Troubleshooting GenBank Submissions: Coding Region Annotation" video up on YouTube

Friday, June 26, 2015

The recording of the June 3rd webinar on troubleshooting internal stop codon errors has been uploaded to [YouTube](#). In this webinar, you will learn how to troubleshoot internal stop codons encountered during coding region (CDS) annotation. You will also learn how to analyze your sequences and uncover problems with BLAST prior to submitting them to GenBank.

For those who cannot access YouTube, the video is also available via [FTP](#). A .txt file of the video's subtitles is also included within that directory.

Subscribe to the [NCBI YouTube channel](#) to be notified of our new videos, which range from quick tips to full webinar presentations.

June 10th webinar "Phylogenetic Trees in Genome Workbench" video up on YouTube

Friday, June 26, 2015

The recording of the June 10th webinar on Genome Workbench's phylogenetic trees has been uploaded to [YouTube](#). In this webinar, you will learn how to use Genome Workbench to work with phylogenetic trees. You will also see the numerous options for importing trees, for visualising, searching and modifying the trees, and for saving or exporting them.

For those who cannot access YouTube, the video is also available via [FTP](#). A .txt file of the video script is also included within that directory.

Subscribe to the [NCBI YouTube channel](#) to be notified of our new videos, which range from quick tips to full webinar presentations.

New dbVar webinar available on NCBI YouTube channel

Thursday, June 25, 2015

A new webinar highlighting [dbVar](#) - NCBI's database of genomic structural variation - is available and consists of two parts: [Part 1](#) is a slide presentation that explores the dbVar website and demonstrates how to navigate its features and tools, including viewing variants in a genome browser; [Part 2](#) is a live demo, with use cases for finding useful information at dbVar as well as direct links to related information at other NCBI resources - for example, clinical assertions at ClinVar.

Subscribe to the [NCBI YouTube channel](#) to be notified of new videos, which range from quick tips to full webinar presentations.

GenBank release 208.0 is now available via FTP

Tuesday, June 23, 2015

[GenBank Release 208.0](#) (6/18/2015) has 185,019,352 non-WGS, non-CON records containing 193,921,042,946 base pairs for sequence data. In addition, there are 258,702,138 WGS records containing 1,038,937,210,221 base pairs of sequence data, as well as 76,974,601 TSA records containing 60,697,472,570 base pairs of sequence data.

During the 66 days between the close dates for GenBank releases 207.0 and 208.0, the non-WGS/non-CON portion of GenBank grew by 69,834,303,408 base pairs and by 14,922,939 sequence records. During that same period, 792,075 records were updated and an average of 54,889 non-WGS/non-CON records were added and/or updated per day.

Between releases 207.0 and 208.0, the WGS component of GenBank grew by 24,303,492,065 base pairs and by 5,163,496 sequence records; the TSA component of GenBank grew by 4,901,140,135 base pairs and by 4,985,013 sequence records.

The total number of sequence data files increased by 59 with this release. The divisions are as follows:

- BCT: 9 new files, now a total of 187
- CON: 2 new files, now a total of 319
- GSS: 4 new files, now a total of 297
- HTG: 9 new files, now a total of 151
- INV: 2 new files, now a total of 128
- PAT: 4 new files, now a total of 223
- PLN: 5 new files, now a total of 112

- PRI: 1 new file, now a total of 49
- TSA: 20 new files, now a total of 195
- VRL: 2 new files, now a total of 36
- VRT: 1 new file, now a total of 46

For downloading purposes, please keep in mind that the uncompressed GenBank flatfiles are approximately 722 GB (sequence files only). The ASN.1 data require approximately 592 GB.



More information about GenBank release 208.0, including details about important changes included in this release, is available in the [release notes](#).

NCBI Southern California Regional Workshops to be held June 30 - July 2

Thursday, June 18, 2015

NCBI Service Desk staff will present a workshop series (June 30 - July 2, 2015) at the Beckman Research Institute of City of Hope in Duarte, California. These [five individual modules](#) include lectures and hands-on demonstrations intended for physicians and researchers who want to learn more about NCBI resources and how to use them to advance research projects.

The workshop series is free, but it is limited to 140 people. Register online at <http://events.signup4.net/UpdateNCBIResource2015>. More information about the event is also available on the registration page and on the flyer below (click [here](#) to download).



Beckman Research Institute of City of Hope
Bioinformatics Core Facility &
Lee Graff Medical and Scientific Library

in collaboration with
The National Center for Biotechnology Information (NCBI)

invites you to a hands-on workshop series presented by
Peter Cooper, Ph.D. and Wayne Matten, Ph.D.
NCBI Service Desk

**“An Update on National Center for
Biotechnology Information (NCBI) Resources”**

June 30 – July 2, 2015
9:00AM - 4:00PM

For additional information, including a detailed schedule
& registration instructions, please visit our website:
<http://events.Signup4.net/UpdateNCBIResource2015>

Registration Deadline: Friday, June 26, 2015
(limited to the first 140 people)

<small>Location: City of Hope Argros Auditorium Arnold and Mabel Beckman Building 1500 E. Duarte Road, Duarte, CA 91010</small>	<small>Contact: Ryan Chiechi Research Finance & Shared Services (626-256-4673 x8900)</small>
---	--

PLEASE NOTE: The workshop is free but registration is required. The sessions will be recorded and made available to those who registered. Each module will have lecture followed by the hands-on computer practice. We encourage people to bring their own laptops. Please arrive 15 mins before each half-day session to allow extra time for computer setup. A limited number (25) of laptops are available for COI attendees on a first come first served basis, determined by the date we receive the online registration.

Light breakfast and lunch will be served.

New YouTube video: "Sequence Viewer: Navigate Objects with Jump Arrows"

Thursday, June 18, 2015

The newest video on the NCBI YouTube channel, "[Sequence Viewer: Navigate Objects with Jump Arrows](#)", introduces jump arrows, a feature recently added to Sequence Viewer. A full list of new features, improvements and fixes can be found in the [release notes](#).

Sequence Viewer is a graphical view of sequences and color-coded annotations on regions of sequences stored in the Nucleotide and Protein databases.

Subscribe to the [NCBI YouTube channel](#) to be notified of our new videos, which range from quick tips to full webinar presentations.

UniVec build 9.0 now available for VecScreen searches and FTP

Tuesday, June 16, 2015

UniVec, NCBI's non-redundant database of vector sequences, has been updated to build 9.0, which enables searches run using NCBI's [VecScreen](#) tool to detect more of the foreign sequences introduced during the cloning or sequencing process. UniVec build 9.0 is also available via [FTP](#).

This build added 252 complete vector sequences and 124 adapter and primer sequences, including many oligonucleotides used in next-generation sequencing protocols, bringing the total number of sequences represented in the UniVec database to 2,658.

UniVec is a non-redundant database of sequences commonly attached to cDNA or genomic DNA during the cloning process. UniVec primarily consists of the unique segment from a large number of vectors but also includes many linker, adapter and primer sequences. Redundant sub-sequences have been eliminated from the database to make searches more efficient and to simplify interpretation of the results. For more details, see the [UniVec page](#).

BLAST+ stand-alone updated to version 2.2.31

Tuesday, June 16, 2015

A new version (2.2.31) of the stand-alone BLAST executables (Linux, Windows and MacOSX on [FTP](#)) is now available. New features include support for BLAST-XML2 specification (information [here](#)) and JSON BLAST output format, as well as several bug fixes and improvements. The BLAST AMI at AWS will also be updated to 2.2.31 (see this [BLAST Help page](#) for more [information](#)). For a full list of improvements, see the [release notes](#).

Related NCBI News stories:

- [June 26, 2014](#): BLAST machine image hosted at Amazon Web Services (AWS)
- [October 16, 2014](#): Amazon Web Services (AWS) Marketplace provides the easiest way to start an NCBI BLAST instance

Complete MERS coronavirus genomes from China and South Korea are in GenBank

Wednesday, June 10, 2015

Two complete MERS coronavirus genomes are in GenBank: one from China ([KT006149](#)) and one from South Korea ([KT029139](#)). In addition, the [MERS coronavirus resource page](#) gives users an easy way to find all sequences related to this pathogen.

The MERS coronavirus resource page has three components designed to support users' discovery activities: the database, the reference genome graphical display, and links to other virus data resources, both external (e.g., CDC, WHO, and HealthMap) and within NCBI.

The database allows you to search for nucleotide and protein sequences by a variety of criteria, including host, sequence patterns, region or country of isolation, and collection or release dates. Using the database, you can:

- Quickly find the sequences you need, through an intuitive search interface, using standardized protein/gene names and metadata
- Select the latest sequences based on date criteria or sorting of results
- Download sequences in many formats and find links to sequences in NCBI databases.

dbSNP build 144 now available

Tuesday, June 09, 2015

dbSNP build 144 data is now available through the integrated NCBI Entrez system and [FTP](#). This release includes data for human, chicken, soybean and horse. Build 144 provides more than 601 million submitted and 191 million reference variants for 4 species. To see complete build statistics, visit the [SNP summary page](#).

NCBI Sequence Viewer version 3.8 available

Wednesday, June 03, 2015

[NCBI Sequence Viewer](#) has recently been updated and now has a new track navigation GUI that allows you to quickly find features like gene, exon, SNP and alignment; it also has a new API option for negative graph values and improved variation tooltips. A full list of new features, improvements and fixes is included in the [release notes](#).

Sequence Viewer is a graphical view of sequences and color-coded annotations on regions of sequences stored in the [Nucleotide](#) and [Protein](#) databases.

June 10th webinar: "Working with Phylogenetic Trees in Genome Workbench"

Wednesday, June 03, 2015

On June 10th, NCBI staff will show you how to use NCBI's powerful Genome Workbench application to work with phylogenetic trees. You will learn about the many options for importing trees, for visualizing, searching and modifying the trees and for saving or exporting them. We will also answer questions and welcome feedback from our participants on future directions for Genome Workbench.

Click [here](#) to sign up for this webinar. Like all our webinars, this will be posted on the NCBI YouTube account after the live presentation; you can subscribe to [our YouTube channel](#) to be notified of all our new videos.

To see upcoming webinars, as well as related materials and recordings from past webinars, please see the [NCBI Webinars page](#).

Conserved Domain Database (CDD) version 3.14 now available online and via FTP

Tuesday, June 02, 2015

Conserved Domain Database (CDD) version 3.14 is now available with 560 new or updated NCBI-curated domains and 50,648 total domain models from CDD's database providers: Pfam, SMART, COG, TIGRFAMs, Protein Clusters, and the NCBI in-house curation project.

You can access CDD at the [Conserved Domains homepage](#) and find updated content on the [CDD FTP site](#). You can also learn about the Conserved Domain Database, how it works and is maintained, and its future in the [most recent Nucleic Acids Research database issue](#).

The SRA Submission App on BaseSpace lets you submit directly to SRA

Tuesday, June 02, 2015

If you use Illumina for next-gen sequencing and want or need to share your genomic data by putting it into a public repository, you can now submit directly to SRA through [BaseSpace](#). You can also submit directly to SRA if you use [Mothur](#) for 16S assembly. Note: it is possible to port data from SRA into BaseSpace; click [here](#) for instructions.

New NCBI YouTube video: "NCBI Minute: Prokaryotic Genome Annotation Update"

Monday, June 01, 2015

The [newest video](#) on the NCBI YouTube channel describes the updates recently made to our prokaryotic genome annotation process. In addition to describing the improvements to RefSeq bacterial and archaeal genome annotation and management, we also provide tips on adapting your workflow and show you how to find more information and help.

Subscribe to the [NCBI YouTube channel](#) to be notified of our new videos, which range from quick tips to full webinar presentations.