NCBI News, January 2017

Multiple Sequence Alignment Viewer 1.3 is now available

Tuesday, January 31, 2017

The new version of the Multiple Sequence Alignment Viewer (MSA Viewer) has implemented a new coloration method and improved tooltips. A full list of new features, improvements, and bug fixes is available in the MSA Viewer release notes.

February 8th NCBI Minute: Finding Gene, Protein and Chemical Names, Aliases and Synonyms

Tuesday, January 31, 2017

Next Wednesday, February 8th, NCBI staff will discuss the systems in the NCBI Gene and PubChem resources that identify and correlate various names used for genes, proteins and chemicals.

Date and time: Wednesday, February 8, 2017 12:00 PM - 12:30 PM EST

Registration URL: https://attendee.gotowebinar.com/register/6498213056303481858

After registering, you will receive a confirmation email with information about attending the webinar. After the live presentation, the webinar will be uploaded to the NCBI YouTube channel. Any related materials will be accessible on the Webinars and Courses page; you can also learn about future webinars on this page.

BLAST+ 2.6.0 offers improved support for accession.version

Thursday, January 26, 2017

The newest version of the BLAST+ executables provides improved support for use of accession.version as the primary identifier, as well as improved speed of blastdbcmd when dumping information from a database.

A number of other bug fixes and improvements are also included in this release. For more information about BLAST+ 2.6.0, please see the release notes.

NLM Citation: NCBI News [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 1991-2012. NCBI News, January 2017.

New NCBI Insights post: Visualize and Interpret Alignment Data with the Multiple Sequence Alignment Viewer

Wednesday, January 25, 2017

The latest post on the NCBI Insights blog introduces the Multiple Sequence Alignment Viewer (MSAV), a resource for visualizing and interpreting alignments for nucleotide and amino acid sequences. The viewer is easily embedded in web pages, readily customizable, and displays alignment data from many sources. Read on at NCBI Insights.

NCBI Insights is the official NCBI blog, where we share science feature stories, quick tips and what's new at NCBI.

January 31st NCBI Minute: New version of E-utilities supports accession.version

Monday, January 23, 2017

Next Tuesday, January 31, 2017, NCBI will present a short webinar that describes and demonstrates new functionality recently introduced to the E-utilities that supports sequence data retrieval.

Date and time: Tuesday, January 31, 2017 12:00 PM - 12:30 PM EST

Registration URL: https://attendee.gotowebinar.com/register/7530877675754064131

After registering, you will receive a confirmation email with information about attending the webinar. After the live presentation, the webinar will be uploaded to the NCBI YouTube channel. Any related materials will be accessible on the Webinars and Courses page; you can also learn about future webinars on this page.

RefSeq release 80 now available; GI identifiers to be removed in next release (March 2017)

Friday, January 13, 2017

RefSeq release 80 is now accessible online, via FTP and through NCBI's programming utilities. This full release incorporates genomic, transcript, and protein data available as of January 9, 2017 and contains 118,059,547 records, including 78,028,152 proteins, 17,862,608 RNAs, and sequences from 66,224 organisms. The release is provided in several directories as a complete dataset and also as divided by logical groupings.

As announced in March 2016, NCBI has implemented the removal of GI numbers from some presentations of nucleotide and protein sequence records. GI sequence identifiers will be removed from flatfile and FASTA formats in the RefSeq FTP release in March 2017.

RefSeq plans to start a comprehensive reannotation of all prokaryotic genomes in a few weeks, which will be included in its entirety in the May 2017 release.

For more information on RefSeq release 80, please see the release notes.

New videos on YouTube: Clone DB and clone placements

Friday, January 13, 2017

Two new videos on the NCBI YouTube channel demonstrate how to use Clone DB and clone placements to assess and improve genome assemblies.

The first video, *Using Clone Placements to Interpret Genome Assemblies*, shows you how to interpret and improve genome assemblies with clone placement data.

The second video, *Clone DB FTP Files: Content and Uses for Genome Assemblies*, teaches you how to use Clone DB FTP files to not only improve genome assemblies but also detect potential structural variation and place clones based on clone end sequences.

Subscribe to the NCBI YouTube channel to receive alerts about new videos ranging from quick tips to full webinar presentations.

GenBank release 217.0 is available via FTP

Wednesday, January 11, 2017

GenBank release 217.0 (12/15/2016) has 198,565,475 traditional records containing 224,973,060,433 base pairs of sequence data. In addition, there are 395,301,176 WGS records containing 1,817,189,565,845 base pairs of sequence data, 142,094,337 TSA records containing 125,328,824,508 base pairs of sequence data, as well as 1,268,690 TLS records containing 584,697,919 base pairs of sequence data.

During the 65 days between the close dates for GenBank releases 216.0 and 217.0, the traditional portion of GenBank grew by 4,421,745,183 base pairs and by 1,174,784 sequence records. During the same period, 726,256 records were updated at an average of 29,247 traditional records added and/or updated per day.

Between releases 216.0 and 217.0, the WGS component of GenBank grew by 140,951,076,595 base pairs and by 32,087,861 sequence records. The TSA component of GenBank grew by 12,119,598,746 base pairs and by 17,894,740 sequence records.

The total number of sequence data files increased by57 with this release. The divisions are as follows:

- BCT: 23 new files, now a total of 304
- CON: 5 new files, now a total of 357
- ENV: 1 new file, now a total of 94
- GSS: 2 new files, now a total of 303
- HTG: 2 new files, now a total of 154

- MAM: 2 new files, now a total of 39
- PAT: 12 new files, now a total of 280
- PLN: 2 new files, now a total of 137
- VRL: 1 new file, now a total of 45

For downloading purposes, please keep in mind that the uncompressed GenBank Release 216.0 flatfiles require approximately 809 GB (sequence files only); the ASN.1 data require approximately 671 GB.

More information about GenBank release 217.0 is available in the release notes, as well as in the README files in the genbank (ftp.ncbi.nih.gov) and ASN.1 (ncbi-asn1) directories.

Genome Workbench 2.11.7 now available

Wednesday, January 04, 2017

The latest version of Genome Workbench includes a number of new features, fixes and improvements like a critical improvement in HTTPS protocol communication with NCBI, improved rendering for translation discrepancies, and improved handling of tracks.

For a full list of changes, please see the Genome Workbench release notes.