NCBI News, December 2015

NCBI staff will attend the International Plant and Animal Genome Conference XXIV in January

Wednesday, December 30, 2015

From January 9-13, 2016, NCBI staff will present posters, give lectures and presentations, and lead workshops at the International Plant and Animal Genome Conference. In addition, NCBI will have a booth (#618) where we will answer any questions you may have; we also welcome your suggestions and comments.

Selected talks and posters:

- **Plenary Lecture:** ***The Genome Era at NCBI – Are We There Yet?*** by Dr. James M. Ostell, Chief of the Information Engineering Branch (IEB) of NCBI (Tuesday, January 12, 2016 8:45 AM)
- **NCBI Genome Resources Workshop** (Monday, January 11, 2016 12:50 PM – 3:00 PM)
- P0294*dbSNP and dbVar: NCBI Databases of Simple and Structural Variations for All Organisms* (Monday, January 11, 2016 10:00 AM – 11:30 AM)
- P0295*Curation of Genome Assemblies by NCBI Tools and Resources* (Monday, January 11, 2016 3:00 PM – 4:30 PM)

To see the complete schedule of NCBI's activities at PAG XXIV, see the Conferences and Presentations page or visit **Booth #618** while at PAG.

BLAST+ executables 2.3.0 now available

Tuesday, December 29, 2015

A new version (2.3.0) of the stand-alone BLAST executables (Linux, Windows and MacOSX available on FTP) is now available. The BLAST VM at cloud providers will be updated soon as well.

This new version includes a beta release of SAM format, as well as support for single-file mode for BLAST XML2 and JSON formats; single-file mode means that results are delivered as a single-file regardless of the number of queries. A number of other

NLM Citation: NCBI News [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 1991-2012. NCBI News, December 2015.

improvements and bug fixes, including a new versioning policy, are included. Please refer to the release notes for the full list of improvements, bug fixes, and more.

GenBank release 211.0 is now available via FTP

Thursday, December 24, 2015

GenBank release 211.0 (12/18/2015) has 189,232,925 non-WGS, non-CON records containing 203,939,111,071 base pairs of sequence data. In addition, there are 317,122,157 WGS records containing 1,297,865,618,365 base pairs of sequence data, as well as 87,488,539 TSA records containing 77,583,339,176 base pairs of sequence data.

During the 60 days between the close dates for GenBank releases 210.0 and 211.0, the traditional (i.e., non-WGS, non-CON) portion of GenBank grew by 1,702,029,512 base pairs and by 860,908 sequence records. During that same period, 1,626,191 records were updated and an average of 41,451 traditional records were added and/or updated each day.

Between releases 210.0 and 211.0, the WGS component of GenBank grew by 75,230,350,867 base pairs and by 7,923,214 sequence records; the TSA component of GenBank grew by 6,666,166,232 base pairs and by 5,698,508 sequence records.

The total number of sequence data files increased by 29 with this release. The divisions are as follows:

- BCT: 8 new files, now a total of 216
- CON: 3 new files, now a total of 327
- ENV: 2 new files, now a total of 88
- INV: 3 new files, now a total of 135
- PAT: 7 new files, now a total of 242
- PLN: 3 new files, now a total of 122
- PRI: 1 new file, now a total of 50
- ROD: 1 new file, now a total of 32
- TSA: 1 less file, now a total of 194
- VRL: 1 new file, now a total of 39
- VRT: 1 new file, now a total of 60

For downloading purposes, please keep in mind that the uncompressed GenBank flat files require approximately 749 GB (sequence files only); the ASN.1 data require approximately 613 GB. More information about GenBank release 211.0 is available in the release notes.

January 7th: Explore new graphical viewer track options with The NCBI Minute

Tuesday, December 22, 2015

On January 7th, NCBI will present a new NCBI Minute webinar, "New track options for getting the most out of NCBI Graphical Viewers". In this webinar, you'll learn how to use these new features to get the most out of Sequence Viewer, Variation Viewer and other NCBI graphical browsers. In addition, we'll show you how to search and quickly find relevant tracks and upload your own custom data.

Date and time: Thursday, January 7, 2016 12:00 – 12:15 PM EST

Registration URL: https://attendee.gotowebinar.com/register/9028024375893233666

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 find information about future webinars on this page.

The NCBI Minute is a series of short webinars that give a brief introduction to a specific topic or NCBI tool.

January 5th webinar: Eukaryotic Genome Data Curation at NCBI

Monday, December 21, 2015

On January 5th, three RefSeq (Reference Sequence) curators will focus on aspects of data curation for eukaryotic organisms. We'll discuss several aspects of manual curation, including sequence analysis, functional annotation, nomenclature review, data validation and community collaboration. We will also highlight how these curation efforts improve the programmatic approaches used by the genome annotation pipelines, which allow NCBI to handle the ever-increasing amount of data generated by researchers.

Date and time: January 5, 2016 1:00 – 2:00PM EST

Registration URL: https://attendee.gotowebinar.com/register/5281805741351150082

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 find information about future webinars on this page.

New on NCBI Insights blog: "The NCBI Minute: quick introductions to NCBI resources"

Thursday, December 17, 2015

Today's post on NCBI Insights discusses the NCBI Minute, a series of short webinars that introduce a new NCBI tool or resource or provide quick tips for using a popular NCBI resource.

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

dbSNP Build 146 for non-human organisms is now available

Wednesday, December 16, 2015

dbSNP Build 146 is accessible via Entrez and FTP. This release includes data for corn, cow, dog, rat, chickpea and rapeseed. Build 146 provides over 380 million submitted variants and 166 million reference variants for 6 species. To see complete build statistics, visit the SNP summary page.

dbSNP, the NCBI Short Genetic Variations database, catalogs short variations in nucleotide sequences from a wide range of organisms.

New NCBI Insights blog posts highlight SRA Toolkit, Run Selector

Friday, December 11, 2015

Today, two new blog posts on NCBI Insights present SRA Toolkit and Run Selector, which allow you to integrate downloaded data sets into pipelines and fine-tune web-based search results, respectively.

"SRA Toolkit: the SRA database at your fingertips"

"SRA Toolkit: the SRA database at your fingertips" briefly explains where to download the SRA Toolkit (on its own webpage, or on GitHub) and describes the various capabilities of the toolkit's command-line executables, which include:

- Streaming data from the NCBI/SRA servers
- Working with restricted-access data from dbGaP (***after*** applying for and receiving access to this data)

Click over to this post on NCBI Insights to learn more ways to use the SRA Toolkit.

"Fine-tune your web-based search results with SRA Run Selector"

This blog post introduces Run Selector, a feature within web-based SRA search that lets you use fields to quickly filter search results to include only data relevant to you. Run Selector also makes it easy to download data or accession lists; this and more is explained in the blog post.

The Sequence Read Archive (SRA) is NCBI's largest growing repository of molecular data. It archives raw sequencing data and alignment information from high-throughput sequencing platforms.

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

December 17th webinar: "Accessing 1000 Genomes Project Data"

Thursday, December 03, 2015

On Thursday December 17th, 2015, NCBI staff will demonstrate how to access 1000 Genomes data through SRA, dbVar, SNP and BioProject, as well as through tracks on annotated human sequences in the graphical sequence viewer and Variation Viewer. Attendees will also learn how to display, search and download individual and genotype level data through the dedicated 1000 Genomes Browser that allows searching by chromosomal position, gene names and other genome markers.

Date and time: Dec 17, 2015 1:00 – 2:00 PM EST

Registration URL: https://attendee.gotowebinar.com/register/5168155820927556866

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 find information about future webinars on this page.

Registration open for December 16th NCBI Minute: "New Faceted Advanced Search in dbGaP Provides Easy Access to Relevant Data"

Thursday, December 03, 2015

On December 16th, the NCBI Minute short webinar will introduce dbGaP's new faceted advanced search interface and show attendees how to use the new interface to easily find data by study, variables, documents and genotypes.

Date and time: Dec 16, 2015, 12:00-12:15 PM EST

Registration URL: https://attendee.gotowebinar.com/register/7869339230869750529

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 find information about future webinars on this page.

The NCBI Minute is a series of short webinars that give a brief introduction to a specific topic or NCBI tool.