From Genes to Genomes – New Features and Data Access

Kim D. Pruitt

NCBI Genome Resources Workshop

PAG XXIII January 10-14, 2015

pruitt@ncbi.nlm.nih.gov

Overview

Poster P1105

- NCBI Genomes – the big picture
- Genome
- Assembly
- Gene
- Finding NCBI News

NCBI Genomes in a Nutshell

Submitter

Sequence → Assembly (Annotate) → Submit

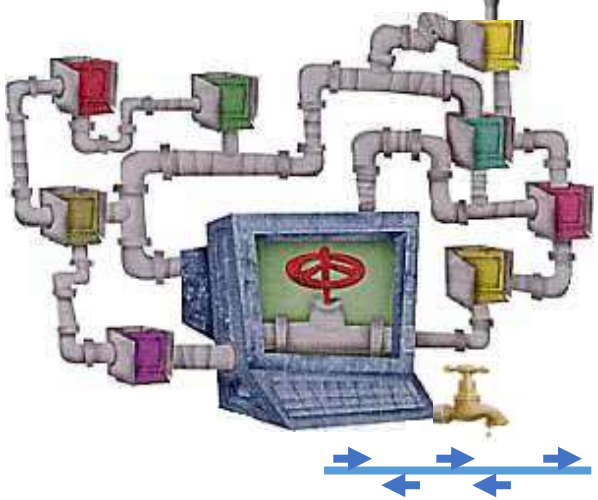
Poster P1112

GenBank/INSDC Genome

Nucleotide	Protein	SRA (reads)	**Sequence**
Assembly	BioSample	BioProject	**Meta-data**
BLAST	FTP	Web eUtils	**Access**

acgcatgttacgca
tagaacactag
cctagaatccnnt

Poster P0051



Derivative Resources

RefSeq	Gene	BLAST
Genome	Tracks	FTP
Reports	Assembly	HomoloGene

RefSeq
Annotation Pipeline
RefSeq Curation

Genome

www.ncbi.nlm.nih.gov/genome/

NCBI Resources How To Sign in to NCBI

Genome Genome Search

Limits Advanced Help

Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

Using Genome
- Help
- Browse by Organism
- Download / FTP
- Download FAQ
- Submit a genome

Custom resources
- Human Genome
- Microbes
- Organelles
- Viruses
- Prokaryotic reference genomes

Other Resources
- Assembly
- BioProject
- BioSample
- Map Viewer
- Protein Clusters

Genome Tools
- BLAST the Human Genome
- Microbial Nucleotide BLAST
- TaxPlot (3-way Genome Comparison)

Genome Annotation and Analysis
- Eukaryotic Genome Annotation
- Prokaryotic Genome Annotation
- PASC (Pairwise Sequence Comparison)

External Resources
- GOLD - Genomes Online Database
- Ensembl Genome Browser
- Bacteria Genomes at Sanger
- Large-Scale Genome Sequencing (NHGRI)

Genome

A central portal to access organism-specific genome information.

- Browse
- Search
- Annotation statistics
- Navigate to related data
- Navigate to Annotation Reports

Genome

NCBI Resources How To Sign in to NCBI

Genome | Genome | rice[orgn] | Search

Save search Limits Advanced Help

Display Settings: Overview Send to:

Tools

BLAST Genome

★New feature!

Oryza sativa (rice)

Reference genome: Oryza sativa Japonica Group (assembly Build 4.0)

Download sequences in FASTA format for **genome**, **transcript**, **protein**

Download genome annotation in **GFF**, **GenBank** or **tablular** format

BLAST against Oryza sativa **genome**

All 6 genomes for species:

Browse the **list**

Download sequence and annotation from **RefSeq** or **GenBank**

Organism Overview ; Genome Assembly and Annotation report [6] ; Plasmid Annotation Report [2] ; Organelle Annotation Report [4]

Oryza sativa (rice)

Oryza sativa overview

Lineage: Eukaryota[1170]; Viridiplantae[146]; Streptophyta[136]; Embryophyta[135]; Trache[134]; Spermatophyta[133]; Magnoliophyta[131]; Liliopsida[31]; Poales[24]; Poaceae[24]; BE clade[21]; Ehrhartoideae[15]; Oryzeae[15]; Oryza[13]; Oryza sativa[1]

Rice is one of the most important food crops in the world and feeds more people than any ot crop. Rice belongs to the genus *Oryza* which includes approximately 24 species. They are w distributed growing in different habitats and different soil types. They show differences in pla growth, yield, pest and disease resistance, stress tolerance More...

Sequence data: genome assemblies: 6; sequence reads: 29 (See Genome Assembly and Annotati

Publications

Representative (genome information for reference and representative genomes)

Reference genome:

Oryza sativa Japonica Group

Submitter: National Institute of Agrobiological Sciences, Japan

Loc	Type	Name	RefSeq	INSDC	Size (Mb)	GC%	Protein	rRNA	tRNA	Gene	Pseudogene
Nuc	Chr	1	NC_008394.4	AP008207.2	45.06	44.0	4,048	-	101	4,279	-
Nuc	Chr	2	NC_008395.2	AP008208.2	36.82	43.5	3,196	-	77	3,411	-
Nuc	Chr	3	NC_008396.2	AP008209.2	37.26	43.8	3,505	-	82	3,710	-
Nuc	Chr	4	NC_008397.2	AP008210.2	35.86	44.3	2,517	-	80	2,704	-
Nuc	Chr	5	NC_008398.2	AP008211.2	30.04	44.0	2,282	-	60	2,398	-
Nuc	Chr	6	NC_008399.2	AP008212.2	32.12	43.6	2,293	-	40	2,436	-
Nuc	Chr	7	NC_008400.2	AP008213.2	30.36	43.6	2,175	-	44	2,308	-
Nuc	Chr	8	NC_008401.2	AP008214.2	28.53	43.4	1,924	-	52	2,067	-
Nuc	Chr	9	NC_008402.2	AP008215.2	23.84	43.7	1,582	-	40	1,690	-
Nuc	Chr	10	NC_008403.2	AP008216.2	23.66	43.8	1,544	-	66	1,673	-
Nuc	Chr	11	NC_008404.2	AP008217.2	30.83	43.4	1,725	-	33	1,843	-
Nuc	Chr	12	NC_008405.2	AP008218.2	27.76	43.0	1,601	-	71	1,776	-
Plastid	Chr	Pltd	NC_001320.1	X15901.1	0.134525	39.0	108	8	38	159	5
MT	Chr	MT	NC_011033.1	BA000029.3	0.49052	43.9	53	3	22	81	6
MT	Plsm	B1	NC_001751.1	D00293.1	0.002135	44.2	2	-	-	-	-

Genome data

Annotation reports

BLAST

Downloads

Organism info.

www.ncbi.nlm.nih.gov/genome/?term=rice

Browse Genome

Genome | Genome | Search

Genome Information by organism

Ruminantia (taxid:9845) | Search by organism | Clear | Download Reports from FTP site

Overview [11] | Eukaryotes [17] | Prokaryotes [0] | Viruses [0] | Plasmids [0]

Download selected records

Items 1 - 11 of 11 << First < Prev Page 1 of 1 Next > Last >>

Organism/Name	Kingdom	Group	SubGroup	Size (Mb)	Chr	Organelles	Plasmids	Assemblies
	All	All	All					
Bison bison	Eukaryota	Animals	Mammals	2828.03	-	1	-	1
Bos indicus	Euk							
Bos mutus	Euk							
Bos taurus	Euk							
Bubalus bubalis	Euk							
Capra aegagrus	Euk							
Capra hircus	Euk							
Capreolus capreolus	Euk							
Odocoileus virginianus	Euk							
Ovis aries	Euk							
Pantholops hodgsonii	Euk							

Overview [11] | Eukaryotes [17] | Prokaryotes [0] | Viruses [0] | Plasmids [0]

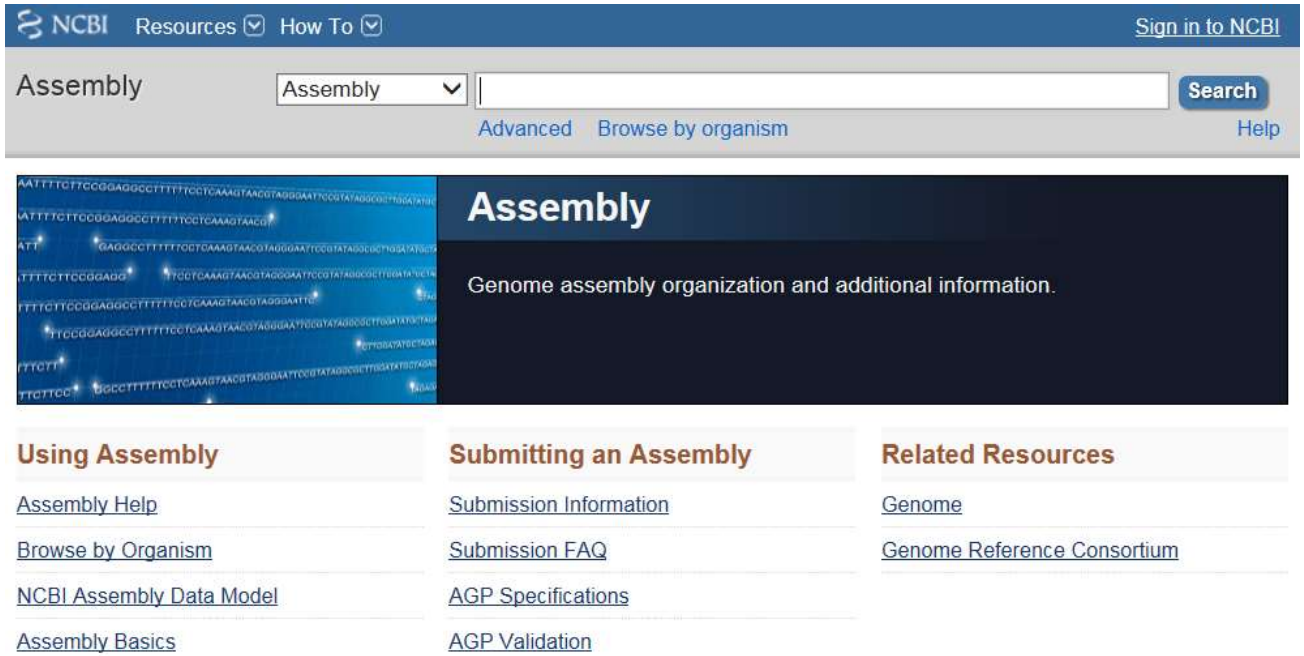
Levels: All, Complete, Chromosome, Scaffold, Contig

Download selected records

Items 1 - 17 of 17 << First < Prev Page 1 of 1 Next > Last >>

Organism/Name	BioSample	BioProject	Group	SubGroup	Assembly	Level	Size (Mb)	GC%	Chrs	Organelle
			-- All E	-- All Euka						
Bison bison bison	SAMN02947321	PRJNA266339	Animals	Mammals	GCA_000754665.1		2828.03	42.20	-	1
Bos grunniens mutus	SAMN00744358	PRJNA221623	Animals	Mammals	GCA_000298355.1		2645.16	42.00	-	1
Bos indicus	SAMN02953802	PRJNA72827	Animals	Mammals	GCA_000247795.2		2673.95	42.30	31	-
Bos taurus	SAMN02898106	PRJNA33843	Animals	Mammals	GCA_000003055.4		2670.04	41.89	30	1
Bos taurus	SAMN02898106	PRJNA33843	Animals	Mammals	GCA_000003055.5		2670.14	41.89	30	1
Bos taurus	SAMN02898106	PRJNA13366	Animals	Mammals	GCA_000003205.4		2983.31	42.30	31	-
Bubalus bubalis	SAMN02191888	PRJNA223666	Animals	Mammals	GCA_000471725.1		2836.17	42.20	-	1
Bubalus bubalis	SAMN00004269	PRJNA40113	Animals	Mammals	GCA_000180995.2		654.914	42.30	-	-
Capra aegagrus	-	PRJEB3140	Animals	Mammals	GCA_000765075.1		2583.32	42.00	-	-

Assembly www.ncbi.nlm.nih.gov/assembly/



Assembly data

A resource that tracks assemblies, meta-data, history, and statistics.

- Accession number
- Update History
- RefSeq reference genomes
- FTP reports
- FTP download
- Statistics
- Submitter
- Sequences

Search Assembly

Taxonomic name
Common name
Assembly name
Submitter

NCBI Resources How To Sign in to NCBI

Assembly | Assembly | Galliformes | Search

Save search Advanced Browse by organism Help

Display Settings: Summary, 20 per page, Sorted by Default Send to:

Results: 6

Breed
Cultivar
Ecotype
Strain

1. Gallus_gallus-4.0
   Organism: Gallus gallus
   Infraspecific name: Breed: Red Jungle fowl, inbred line UCD001
   Sex: female
   Submitter: International Chicken Genome Consortium
   Date: 2011/11/22
   Assembly level: Chromosome
   Genome representation: full
   RefSeq category: representative genome
   Synonyms: galGal4
   GenBank assembly accession: GCA_000002315.2 (latest)
   RefSeq assembly accession: GCF_000002315.3 (latest)
   IDs: 317958 [UID] 317448 [GenBank] 317958 [RefSeq]

Reference genome
Representative genome

2. Turkey_5.0
   Organism: Meleagris gallopavo
   Infraspecific name: Breed: Aviagen turkey brand Nicholas breeding stock

Filter your results:

All (10)
Latest (6)
Representatives (5)

Manage Filters

Find related data

Search details

Recent activity

Improvements

Enhanced search

Infraspecific names

RefSeq category

NCBI Resources How To Sign in to NCBI

Assembly

Assembly Search

Advanced Browse by organism Help

Display Settings: Full Report

Send to:

Turkey_5.0

Organism name: Meleagris gallopavo

Infraspecific name: Breed: Aviagen turkey brand Nicholas breeding stock

Isolate: NT-WF06-2002-E0010

Sex: female

BioSample: SAMN02981253

Submitter: Turkey Genome Consortium

Date: 2014/11/24

Assembly level: Chromosome

Genome representation: full

RefSeq category: representative genome

GenBank assembly accession: GCA_000146605.3 (latest)

RefSeq assembly accession: GCF_000146605.2 (latest)

RefSeq assembly and GenBank assembly identical: yes

WGS Project: ADDD02

Assembly method: MaSuRCA v. 1.9.2

Expected final version: Yes

Genome coverage: 35.0x

Sequencing technology: Illumina GAII; Sanger; 454

See Genome Information for **Meleagris gallopavo**

Access the data

BLAST search the assembly

Download the full sequence report

Download the statistics report

GenBank FTP site

RefSeq FTP site

New!

Assembly Information

Assembly Help

Assembly Basics

NCBI Assembly Data Model

Related Information

BioProject

BioSample

Genome

Nucleotide INSDC

Nucleotide RefSeq

NIH www.ncbi.nlm.nih.gov/assembly/GCF_000146605.2

NCBI Resources How To Sign in to NCBI

Assembly | Assembly | Search

Advanced Browse by organism Help

Display Settings: Full Report

Send to:

Turkey_5.0

Organism name: Meleagris gallopavo

Infraspecific name: Breed: Aviagen turkey brand Nicholas breeding stock

Isolate: NT-WF06-2002-E0010

Sex: female

BioSamp

Submitte

Date: 2

Assembl

Genome

RefSeq c

GenBank

RefSeq a

RefSeq a

WGS Pro

Assembl

Expected

Genome coverage: 35.0x

Sequencing technology: Illumina GAII; Sanger; 454

See Genome Information for **Meleagris gallopavo**

Access the data

BLAST search the assembly

Download the full sequence report

Download the statistics report

GenBank FTP site

RefSeq FTP site

★ New!

Nucleotide INSDC

Nucleotide RefSeq

History (Hide revision history)

GenBank Assembly Accession		RefSeq Assembly Accession	Assembly Name	Assembly Level	Status
GCA_000146605.3	**=**	**GCF_000146605.2**	**Turkey_5.0**	**Chromosome**	**Latest GenBank, Latest RefSeq**
GCA_000146605.2	=	GCF_000146605.1	Turkey_2.01	Chromosome	Replaced GenBank, Replaced RefSeq
GCA_000146605.1	n/a	n/a	Turkey_2.01	Chromosome	Replaced GenBank

www.ncbi.nlm.nih.gov/assembly/GCF_000146605.2

Turkey_5.0 Assembly Statistics

Global statistics

Total sequence length	1,128,339,136
Total assembly gap length	35,294,427
Gaps between scaffolds	2,520
Number of scaffolds	233,806
Scaffold N50	3,801,642
Number of contigs	296,331
Contig N50	26,671
Total number of chromosomes and plasmids	33

www.ncbi.nlm.nih.gov/assembly/GCF_000146605.2

Turkey_5.0 Assembly Statistics

Assembly Definition | Assembly Statistics

Global assembly definition

Download the full sequence report

Click on the table row to see sequence details in the table to the right

Assembly Unit Name
Primary Assembly
non-nuclear

Assembly Unit: Primary Assembly (GCF_000146615.2)

Molecule name	GenBank sequence		RefSeq sequence	Unlocalized sequences count
Chromosome 1	CM000962.2	=	NC_015011.2	2,000
Chromosome 2	CM000963.2	=	NC_015012.2	973
Chromosome 3	CM000964.2	=	NC_015013.2	604

www.ncbi.nlm.nih.gov/assembly/GCF_000146605.2

Turkey_5.0 Assembly Statistics

Assembly Definition | Assembly Statistics

Assembly statistics

Primary Assembly | non-nuclear

Molecule	Sequence Role	Total Length	Scaffold Count	Ungapped Length	Scaffold N50	Spanned Gaps	Unspanned Gaps
Chromosome 1	**All**						
	Assembled	192,233,180	2,318	189,363,663	3,494,005	10,740	317
	molecule	190,651,702	318	187,891,158	3,707,307	10,690	317
	Unlocalized scaffolds	1,581,478	2,000	1,472,505	825	50	0

www.ncbi.nlm.nih.gov/assembly/GCF_000146605.2

Download (Genomes FTP redesign) New!

Genome FTP FAQ: www.ncbi.nlm.nih.gov/genome/doc/ftpfaq/

Advantages:

- Comprehensive provision of GenBank and RefSeq genomes data available in NCBI's Assembly database
- Provides a consistent set of files including:
  - FASTA
  - GenBank/GenPept
  - GFF3
  - Md5checksums
  - Assembly structure
- **Consistent use of accession.version as the primary sequence ID for both GFF and FASTA files** to facilitate use of these data in some public domain RNAseq read mapping tools.

Download (Genomes FTP redesign)

- Access:
  - Follow links from Assembly or Genome
  - Browse
    - ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/
    - ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/
  - Find assemblies using summary report files and go directly to ftp://ftp.ncbi.nlm.nih.gov/genomes/all/ (don't browse!)

```
README.txt
archaea
assembly_summary_genbank.txt
bacteria
fungi
invertebrate
other
plant
protozoa
vertebrate_mammalian
vertebrate_other
```

```
README.txt
archaea
assembly_summary_refseq.txt
bacteria
fungi
invertebrate
plant
protozoa
vertebrate_mammalian
vertebrate_other
```

Gene resource

www.ncbi.nlm.nih.gov/gene/

NCBI Resources How To Sign in to NCBI

Gene Gene Search

Advanced Help

Filters activated: Current only. Clear all

Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene	Gene Tools	Other Resources
Gene Quick Start	Submit GeneRIFs	HomoloGene
FAQ	Submit Correction	OMIM
Download/FTP	Statistics	RefSeq
RefSeq Mailing List	BLAST	RefSeqGene
Gene News	Genome Workbench	UniGene
Factsheet	Splign	Protein Clusters

Representative queries

Find genes by...	Search text
free text	human muscular dystrophy
chromosome and symbol	(II[chr] OR 2[chr]) AND adh*[sym]
partial name and multiple species	alive[prop] AND transporter[title] AND ("Drosophila melanogaster"[orgn] OR "Mus musculus"[orgn])
associated sequence accession	M11313[accn]

Gene data

Annotation	Nomenclature
Bibliography	Pathways
Gene ID	Publications
Gene Ontology	RefSeq
Homology	Variation

Improved features

- Annotation history
- Homology data
- RNA-Seq tracks
- Search Filters
- Tabular Search Results

Gene resource

www.ncbi.nlm.nih.gov/gene/

NCBI Resources How To Sign in to NCBI

Gene Gene Search

Advanced Help

Filters activated: Current only. Clear all

Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene	Gene Tools	Other Resources
Gene Quick Start	Submit GeneRIFs	HomoloGene

Gene: a gene-centered information resource at NCBI.

Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, Tolstoy I, Tatusova T, **Pruitt** KD, Maglott DR, Murphy TD.
Nucleic Acids Res. 2014 Oct 29. pii: gku1055. [Epub ahead of print]
PMID: 25355515 [PubMed - as supplied by publisher] **Free Article**

Representative queries

Find genes by...	Search text
free text	human muscular dystrophy
chromosome and symbol	(1[chr] OR 2[chr]) AND adh*[sym]
partial name and multiple species	alive[prop] AND transporter[title] AND ("Drosophila melanogaster"[orgn] OR "Mus musculus"[orgn])
associated sequence accession	M11313[accn]

Gene data

Annotation	Nomenclature
Bibliography	Pathways
Gene ID	Publications
Gene Ontology	RefSeq
Homology	Variation

Improved features

- Annotation history
- Homology data
- RNA-Seq tracks
- Search Filters
- Tabular Search Results

Searching Gene

NCBI Resources How To Sign in to NCBI

Gene | Gene | turkey[organism] | Search

Save search Advanced Help

Show additional filters

Clear all

Gene sources
Genomic
Mitochondria
Organelles

Categories
Alternatively spliced
Annotated genes
Non-coding
Protein-coding
Pseudogene

Sequence content
Ensembl
RefSeq

Status clear
✓ Current only

Chromosome locations
Select ...

Clear all

Show additional filters

Display Settings: Tabular, 20 per page, Sorted by Relevance Send to:

Hide sidebar >>

<< First < Prev Page 1 of 1059 Next > Last >>

Results: 1 to 20 of 21170

Filters activated: Current only. Clear all to show 21821 items.

Name/Gene ID	Description	Location	Aliases
ADRB1 ID: 100303680	adrenoceptor beta 1 [*Meleagris gallopavo* (turkey)]	Chromosome 8, NC_015018.2 (28488986..28492113)	ADRB2
LYZ ID: 100547044	lysozyme [*Meleagris gallopavo* (turkey)]	Chromosome 1, NC_015011.2 (34890259..34893996)	
PRL ID: 100303694	prolactin [*Meleagris gallopavo* (turkey)]	Chromosome 3, NC_015013.2 (5357235..5364494)	
GPC1 ID: 693243	glypican 1 [*Meleagris gallopavo* (turkey)]	Chromosome 11, NC_015021.2 (594707..674264)	
ACR ID: 692129	acrosin [*Meleagris gallopavo* (turkey)]	Chromosome 1, NC_015011.2 (123218..129069, complement)	Proacro1

Filters: Manage Filters

Find related data
Database:
Select
Find items

Search details

Recent activity

Improved features

- Filters
- Hide right sidebar
- Tabular results
- Download table

www.ncbi.nlm.nih.gov/gene/

Searching Gene

NCBI Resources How To Sign in to NCBI

Gene | Gene | turkey[organism] | Search

Save search Advanced Help

Show additional filters

Clear all

Gene sources
Genomic
Mitochondria
Organelles

Categories
Alternatively spliced
Annotated genes
Non-coding
Protein-coding
Pseudogene

Sequence content
Ensembl
RefSeq

Status clear
✓ Current only

Chromosome locations
Select ...

Clear all

Show additional filters

Display Settings: Tabular, 20 per page, Sorted by Relevance

Send to:

Hide sidebar >>

Results: 1 to 20 of 21170

Filters activated: Current only. Clear all

Name/Gene ID	Description	Loca	
ADRB1 ID: 100303680	adrenoceptor beta 1 [*Meleagris gallopavo* (turkey)]	Chr NC_ (284	
LYZ ID: 100547044	lysozyme [*Meleagris gallopavo* (turkey)]	Chr NC_ (348	
PRL ID: 100303694	prolactin [*Meleagris gallopavo* (turkey)]	Chr NC_ (535	
GPC1 ID: 693243	glypican 1 [*Meleagris gallopavo* (turkey)]	Chromosome 11, NC_015021.2 (594707..674264)	
ACR ID: 692129	acrosin [*Meleagris gallopavo* (turkey)]	Chromosome 1, NC_015011.2 (123218..129069, complement)	Proacro1

Choose Destination

(•) File ( ) Clipboard
( ) Collections

Download 21170 items.

Format
Tabular (text)

Sort by
Relevance

Relevance
Gene Weight
Name
Chromosome

Create File

Improved features

- Download table
  - NCBI TaxID
  - Organism name
  - GeneID
  - Status
  - Symbol
  - Gene name
  - Chromosome
  - Annotated location
  - Exon count

Search Gene by Assembly! ★New feature!

Find Turkey interleukin* genes annotated in the new RefSeq assembly, that were NOT annotated in the previous RefSeq assembly:

- interleukin AND Turkey_5.0 NOT Turkey_2.01
- interleukin AND GCF_000146605.2 NOT GCF_000146605.1

Gene | interleukin[gene/protein name] AND Turkey_5.0 NOT Turkey_2.01 | Search

Save search Advanced Help

Display Settings: Tabular, 20 per page, Sorted by Relevance Send to:

Results: 1 to 20 of 33 << First < Prev Page 1 of 2 Next > Last >

Filters activated: Current only. Clear all to show 33 items.

Name/Gene ID	Description	Location	Aliase
IL6R ID: 104917170	interleukin 6 receptor [*Meleagris gallopavo* (turkey)]		
IL18RAP ID: 104916911	interleukin 18 receptor accessory protein [*Meleagris gallopavo* (turkey)]	Chromosome 1, NC_015011.2 (130299426..130316105)	

SIDEBAR

No Location information indicates gene was annotated on a scaffold

Improved ortholog reporting

Gene | Gene | turkey[orgn] AND IL18 | Search

Save search Advanced Help

Display Settings: Full Report Send to: Hide sidebar >>

Filters activated: Current only. Clear all to show 1 items.

IL18 interleukin 18 [*Meleagris gallopavo* (turkey)]

Gene ID: 100303691, updated on 31-Dec-2014

Summary

Gene symbol	IL18
Gene description	interleukin 18
See related	Ensembl:ENSMGAG00000004590
Gene type	protein coding
RefSeq status	MODEL
Organism	Meleagris gallopavo
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria; Aves; Neognathae; Galloanserae; Galliformes; Phasianidae; Meleagridinae; Meleagris
Also known as	IL-18
Orthologs	human mouse all

★ New feature!

Table of contents
- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Pathways from BioSystems
- General gene information
  - Homology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links

Orthology Data sources:

- Eukaryotic Annotation Pipeline
- Homologene
- OrthoDB

NIH www.ncbi.nlm.nih.gov/gene/100303691

Annotation on previous assembly

IL18 interleukin 18 [*Meleagris gallopavo* (turkey)]

Genomic context

Location: chromosome: 26

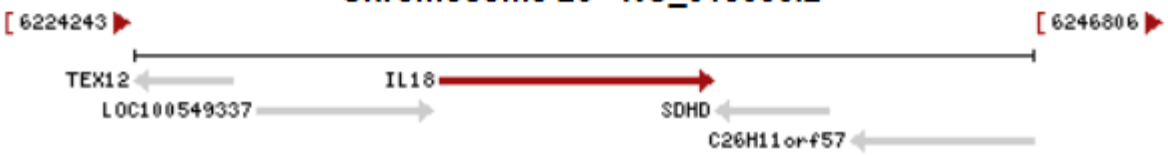
See IL18 in MapViewer

Exon count: 5

Annotation release	Status	Assembly	Chr	Location
101	current	Turkey_5.0 (GCF_000146605.2)	26	NC_015036.2 (6231939..6238766)
Build 1.1	previous assembly	Turkey_2.01 (GCF_000146605.1)	26	NC_015036.1 (6578157..6581114)

New feature!

Chromosome 26 - NC_015036.2



www.ncbi.nlm.nih.gov/gene/100303691

Poster: P1142 (display)

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_015036.2 Chromosome 26 Reference Turkey_5.0 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

Display current OR previous assembly!!

★**New feature!**

Chromosome overview

Tools Configure

Genes, Known RefSeqs, NCBI Meleagris gallopavo Annotation Release 101

Genes, NCBI Meleagris gallopavo Annotation Release 101, 2014-12-18

Annotated RefSeq transcript/protein

RNA-seq exon coverage, aggregate (filtered, log2 scaled), NCBI Meleagris gallopavo Ann.

RNA-seq intron-spanning reads, aggregate (filtered, log2 scaled), NCBI Meleagris gallo.

RNA-seq intron features, aggregate (filtered), NCBI Meleagris gallopavo Annotation Rel.

★**New feature!**

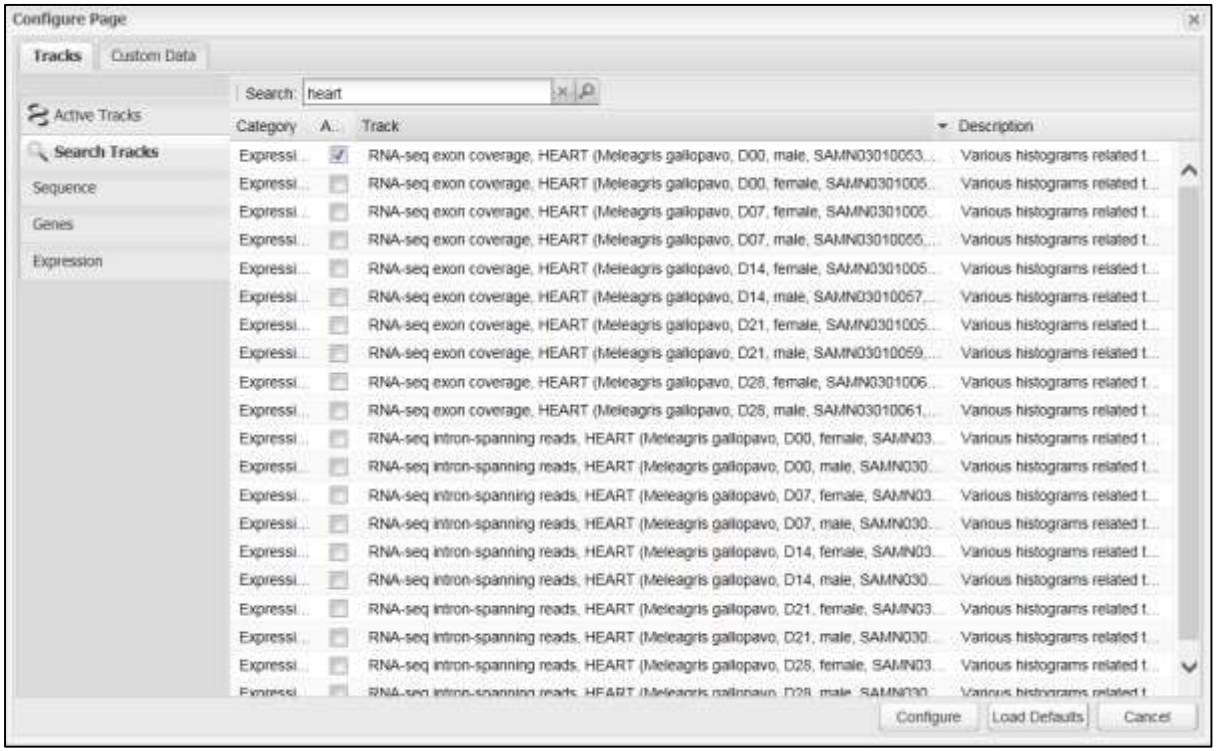
RNA-Seq evidence:

- Histogram
- Reads/intron
- Aggregate and per samples

www.ncbi.nlm.nih.gov/gene/100303691

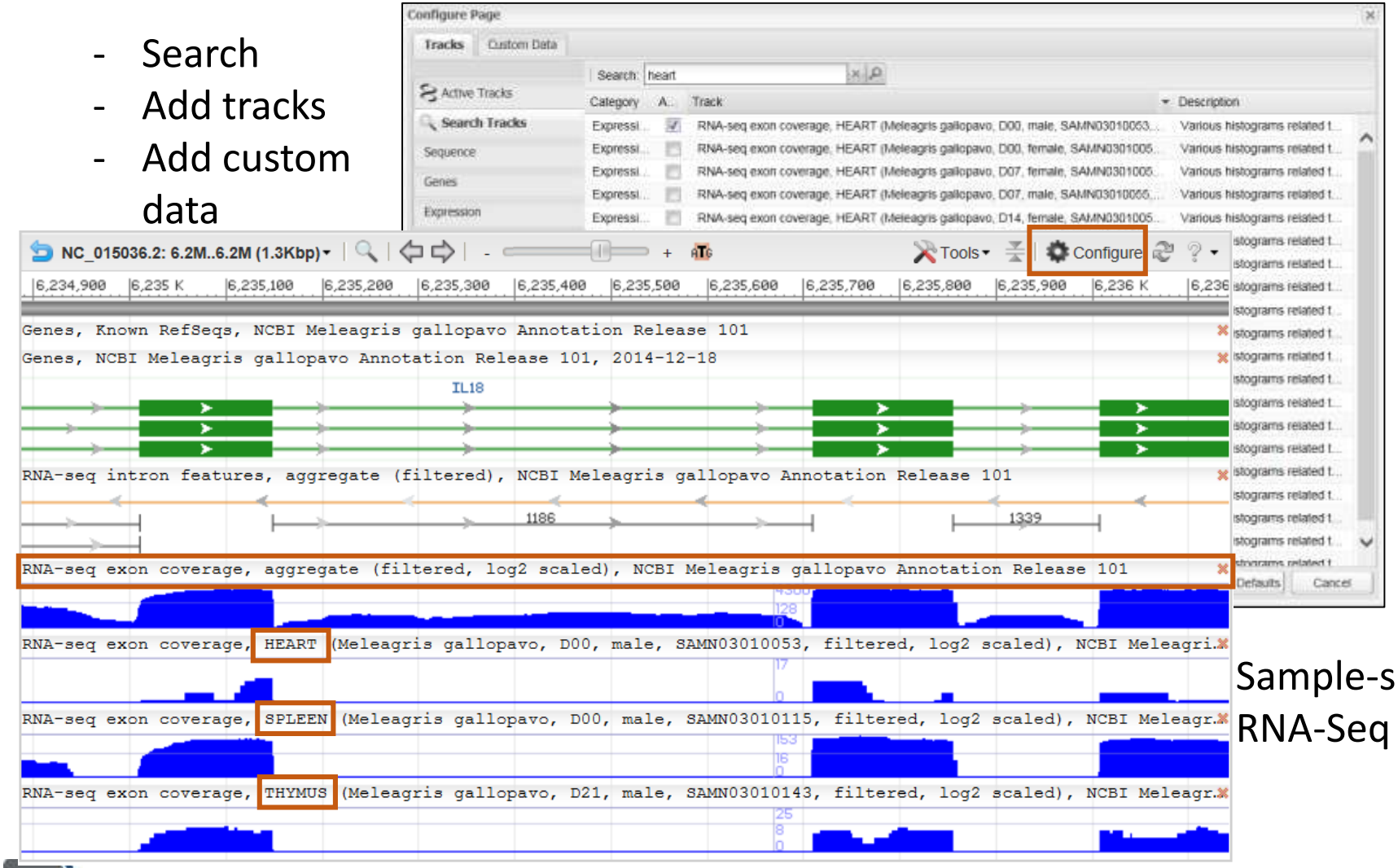
Configure the Gene graphic display

- Search
- Add tracks
- Add custom data



Configure the Gene graphic display

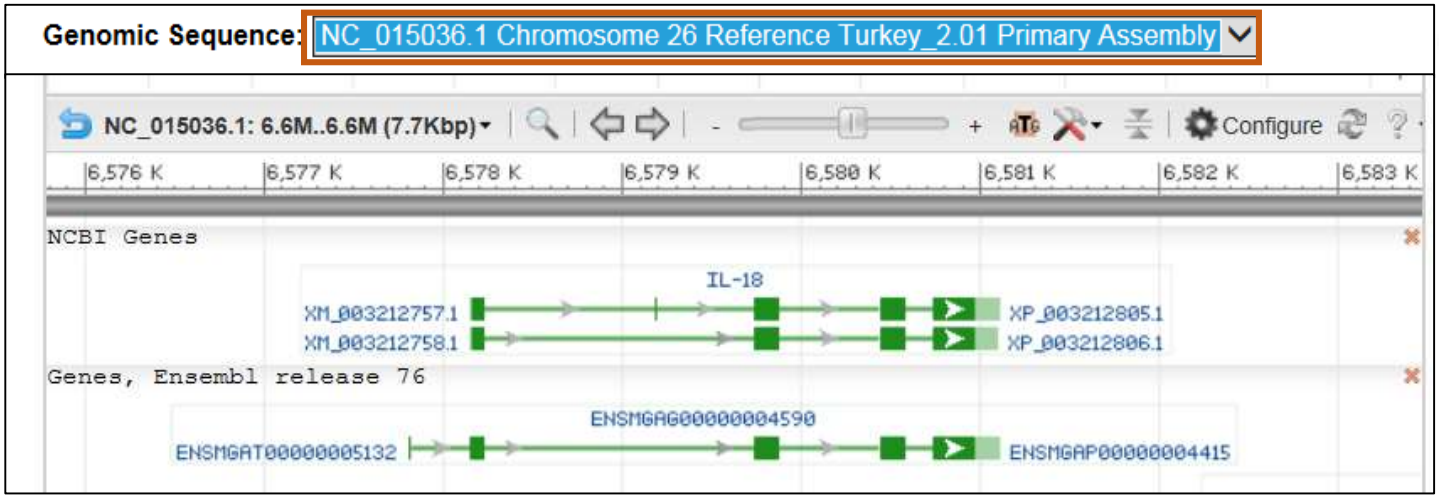
- Search
- Add tracks
- Add custom data



Sample-specific RNA-Seq tracks

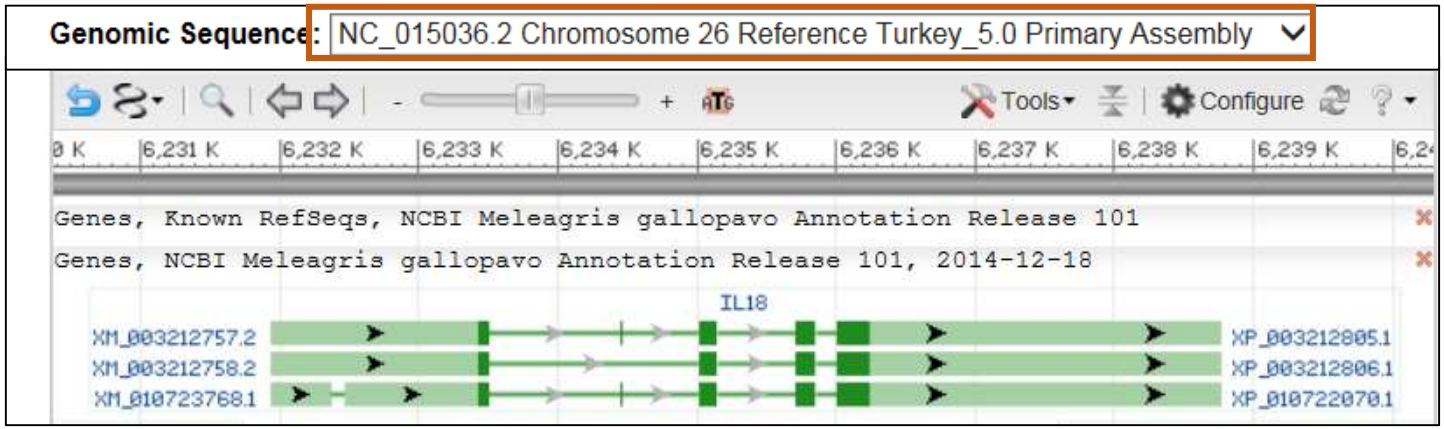
Annotation on previous assembly

IL18 interleukin 18 [*Meleagris gallopavo* (turkey)]



★New feature!

Previous assembly



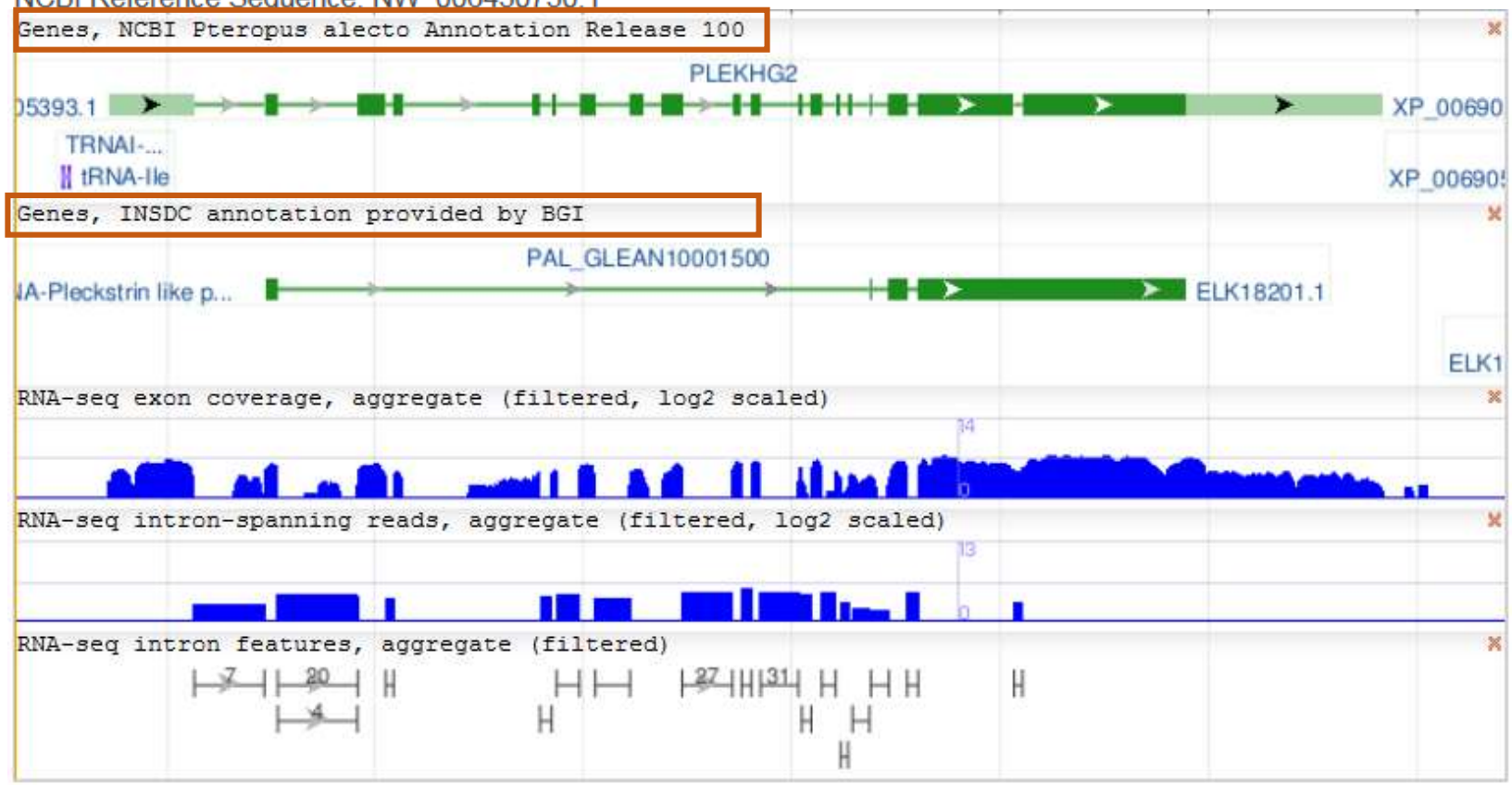
New assembly

Coming soon- GenBank annotation tracks

Pteropus alecto unplaced genomic scaffold, ASM32557v1 scaffold271, whole genome shotgun sequence

Black flying fox (a bat)

NCBI Reference Sequence: NW_006430730.1



Many exons

Few exons

Coming soon- GenBank annotation tracks

Pteropus alecto unplaced genomic scaffold, ASM32557v1 scaffold271, whole genome shotgun sequence

Black flying fox (a bat)

NCBI Reference Sequence: NW_006430730.1

Genes, NCBI Pteropus alecto Annotation Release 100

PLEKHG2

05393.1 XP_00690

TRNAI-...

tRNA-Ile

XP_00690

Many exons

Genes, INSDC annotation provided by BGI

PAL_GLEAN10001500

A-Pleckstrin like p... ELK18201.1

Few exons

RNA-seq exon coverage, aggregate

RNA-seq intron-spanning reads

RNA-seq intron features, aggr

Human PLEKHG2 annotated on GRCh38 (many exons)

Genomic Sequence: NC_000019.10 chromosome 19 reference GRCh38 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

NC_000019.10: 39M..39M (21Kbp) Tools Configure

39,410 K | 39,412 K | 39,414 K | 39,416 K | 39,418 K | 39,420 K | 39,422 K | 39,424 K | 39,426 K | 39,428 K

Genes, NCBI Homo sapiens Annotation Release 106

PLEKHG2

XM_006723330.1 XP_006723393.1

XM_006723334.1 XP_006723397.1

NM_022835.2 NP_073746.

XM_005259163.1 XP_005259220.1

XM_006723331.1 XP_006723394.1

XM_006723332.1 XP_006723395.1

XM_006723333.1 XP_006723396.1

XM_006723329.1 XP_006723392.1

How to find NCBI News

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information

All Databases Search

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | NCBI News

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How Tos: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

Education Resources

Central point of access for help documents, teaching materials, news outlets, and other educational resources.

1 2 3 4 5 6 7 8

Popular Resources

PubMed
Bookshelf
PubMed Central
PubMed Health
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI Announcements

GenBank release 205.0 is now available via FTP
Dec 16, 2014

Release 205.0 (12/12/2014) has

Bald eagle and other bird genome sequence and annotation data publicly available at NCBI
Dec 11, 2014

Citation Exporter Feature Now Available in PubMed Central
Dec 9, 2014

PubMed Central (PMC) has added a

More...

Follow us on

Archives

Year: 2014

Jan	Feb	Mar	Apr
May	Jun	Jul	Aug
Sep	Oct	Nov	Dec

Archives prior to July, 2012

Twitter
Facebook
LinkedIn
RSS
NCBI Announce
NCBI Insights
YouTube

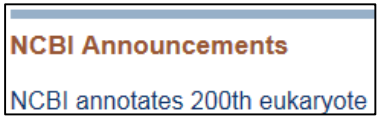
www.ncbi.nlm.nih.gov

Keeping up with NCBI!

NCBI Handbook – resource documentation
www.ncbi.nlm.nih.gov/books/NBK143764/

NCBI YouTube channel – tutorials & seminars
www.youtube.com/user/NCBINLM

NCBI Twitter: twitter.com/NCBI



NCBI News - Articles, resource updates
www.ncbi.nlm.nih.gov/news/

Gene RSS Feed – Announcements, News
www.ncbi.nlm.nih.gov/feed/rss.cgi?ChanKey=genenews

Contact NCBI: info@ncbi.nlm.nih.gov

NCBI Genome Resources Workshop

- 12:50– 1:15 Improving the Flow of Data to National Center of Biotechnology Information Sequence Repositories, SRA and GenBank P1112
  Ilene Mizrachi
- 1:15– 1:40 Variation at NCBI: Resources, Tools and Submissions P1109
  Jennifer Lee
- 1:40– 2:05 Update on the Eukaryotic Genome Annotation Pipeline P0051
  Francoise Thibaud-Nissen
- 2:05– 2:30 From Genes to Genomes – New Features and Data Access P1105
  Kim D. Pruitt
- 2:30– 2:50 Annual Report on Genome Sequencing Projects
  Tatiana Tatusova
- 2:50-3:00 Questions & Discussion

Visit NCBI Booth 618

Contact us: info@ncbi.nlm.nih.gov