

NCBI News, January 2011

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

Created: January 10, 2011; Updated: January 28, 2011.

NCBI Discovery Workshops: Feb 15-16, 2011

NCBI will present a two-day workshop on February 15-16, on the NIH campus in Bethesda, Maryland. The course is free and is open to anyone interested in NCBI resources. The workshops provide hands-on experience exploring practical examples using tools and databases on the NCBI website. The four workshops are Sequences, Genomes, and Maps; Proteins, Domains and Structures; NCBI BLAST Services; and Human Variation and Disease Genes. For more information see the [Discovery Workshop page](#), which also includes a registration link.

Updated Resources for Genomic Libraries and Clones

NCBI has updated resources for finding genomic libraries and genomic clones from genome sequencing projects for a large number of organisms. The new [CloneDB](#) (Figure 1, *Top panel*) replaces the Clone Registry as the resource for finding descriptions, sources, and detailed statistics on available genomic libraries for a large number of organisms.

The new [Library Browser](#) (Figure 1, *Bottom panel*) allows filtering by organism, vector type, distributors, and number of associated database end or insert sequences. The linked [Clone Finder](#) (Figure 2), now available for human, mouse, rat, cow, horse, pig, and zebra finch, quickly identifies clones that span regions on assembled genomes.

The [Clone Finder](#) locates clones by chromosomal position or by features such as genes, SNPs, markers, or transcript sequence accession number. Clones may also be found in regions bounded by any two markers (Figure 2, *Middle panel*). The initial query may be refined to specific mapping data sets, population sources, and libraries. The graphical display in Clone Finder shows features annotated on the genome including assembled contigs, their components, genes, and aligned transcripts (Figure 2, *Bottom panel*).

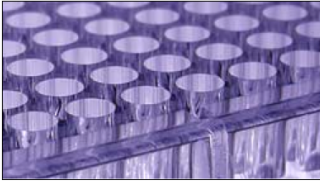
Together CloneDB's [Library browser](#) and the [Clone Finder](#) provide essential access to these important molecular reagents.

New Gene-BioSystems Links Highlight the Gene in the Biological Pathway

Links from [BioSystems](#) are now fully listed in a separate section of Gene records (Figure 3, *Top panel*). Each of these new links point to a specific pathway or system and leads to the record in BioSystems with the specific gene highlighted (Figure 3, *Bottom panel*). For pathways imported from the [Kyoto Encyclopedia of Genes and](#)

CloneDB

Connecting the lab with the genome



Clone DB

Clone DB is a database that integrates information about clones and libraries, including sequence data, map positions and distributor information. It replaces the former NCBI Clone Registry.

Getting Started

- [An overview of Clone DB](#)
- [FAQ](#)
- [News and Announcements](#)

Tools

- [Genomic clone library browser](#)
- [Clone DB Distributors](#)
- [Clone DB FTP site](#)
- [Clone Finder](#)
- [Search Old Clone Registry](#)

Related Resources

- [NCBI MapViewer](#)
- [NHGRI Structural Variation Project](#)
- [Human BAC Resource](#)
- [CCAP Clones](#)

Clone DB: Genomic Clone Library Browser

This browser provides a listing of genomic clone libraries with records in Clone DB. Data filters are provided to assist with navigation of the table.

- [Library Distributors](#): Contact information for distributors in this library browser
- [FAQ](#): Frequently asked questions about this library browser

Filters currently applied

[undo] (organism=Homo sapiens)

[undo] (vector=BAC)

[undo] (distributor=CHORI)

Source: NCB

Genomic Clone Library Browser						
Library Name	Library Abbreviation	Vector types	Distributors	Total clones	Total end sequences	Total insert sequences
RPCI human BAC library 11	RP11	BAC	Empire Genomics imaGenes CHORI Invitrogen	292,384	391,211	32,849
The CHORI-17 BAC Library from a hydatidiform (haploid) mole	CH17	BAC	CHORI	164,737	325,659	246
RPCI human BAC library 13	RP13	BAC	Empire Genomics CHORI	2,778	0	1,128
CHORI-507 Human BAC Library	CH507	BAC	CHORI			
CHORI 15 Human female BAC library	CH15	BAC	CHORI			
CHORI-16 Human sheared BAC library	CH16	BAC	CHORI			
CHORI 502 Human MHC BAC Library	CH502	BAC	CHORI			
CHORI 501 Human MHC BAC Library	CH501	BAC	CHORI			
CHORI 14 Human male BAC library	CH14	BAC	CHORI Empire Genomics			

Filter by End Sequence

>=100

>=1000

>=10,000

>=100,000

>=1,000,000

Source: NCB

Filter by Insert Sequence

>=100

>=500

>=1000

>=5000

>=10,000

Source: NCB

RPCI human BAC library 11

Library Summary

Library Name:	RPCI human BAC library 11
Library Abbreviation:	RP11
Organism:	Homo sapiens
Distributors:	CHORI, Empire Genomics, imaGenes, Invitrogen
Vector type(s):	BAC
# clones Clone DB:	292,384
# end sequences Clone DB:	391,211
# insert sequences Clone DB:	32,849
# clones with both ends sequenced:	132,872

Library Details

DNA Source
Library Construction
Library Statistics

DNA Source

Library segment	Sex	Organ
ALL	male	blood

Figure 1. The CloneDB homepage (Top panel) and the associated Genomic Clone Library Browser (Bottom panel). The Library Browser provides Filters to narrow down the selected libraries. The display has the *Homo sapiens* organism filter and BAC vector filters applied. The inset shows the record for the RP11 BAC library with links to Distributors.

CloneFinder Home

Clone Finder is a tool developed to facilitate the identification of clones within given genomic regions.

Related Resources

- NCBI Home
- NCBI Clone
- NCBI Web Search
- NCBI Site map
- Genome Browser agreement
- Genome Biology
- Taxonomy
- Entrez (Global Query)
- BLAST
- Map Viewer

Vertebrates

Mammals

Primates

Scientific name	Common name	Build
<i>Homo sapiens</i>	human	Build 37.2 Build 36.3

Rodents

Scientific name	Common name	Build
<i>Mus musculus</i>	laboratory mouse	Build 37.1
<i>Rattus norvegicus</i>	rat	RGSC v3.4

Other Mammals

Scientific name	Common name	Build
<i>Bos taurus</i>	cattle	5.2 Btau 4.0 EquCab2.0
<i>Equus caballus</i>	horse	Scrofa9.2
<i>Sus scrofa</i>	pig	Scrofa5

Other Vertebrates

Scientific name	Common name	Build
<i>Gallus gallus</i>	chicken	Build 2.1
<i>Taeniopygia guttata</i>	zebra finch	Build 1.1

Invertebrates

Insects

Scientific name	Common name	Build
<i>Apis mellifera</i>	honey bee	Amel_4.0

Clone Search

Homo sapiens Clone Finder Build: **Build 37.2 (current)** [Change Build](#)

Specify Region

Search by Position | **Search by Feature**

Region	Feature type	Feature name
Chromosome: 19	From: Gene	BCAM
Assembly: GRCh37.p2	To: Gene	SFRS16

[Go](#)

Select Region

Assembly	Chromosome	Begin	End	Length
GRCh37.p2	19	45,312,338	45,594,782	282,445

Select placement ranges to include

Begin	End	Length
<input checked="" type="checkbox"/> 45,312,338	45,324,678	12,341
<input checked="" type="checkbox"/> 45,582,518	45,594,782	12,265
<input type="checkbox"/> 49,298,319	49,314,320	16,002

[Find Clones](#)

Set Data display filters

- Dataset selection
- DNA Source
- Population Selection
- Library Selection Check all Clear all

BAC vectors

<input type="checkbox"/> CH17	<input type="checkbox"/> CTA	<input type="checkbox"/> CTB	<input type="checkbox"/> CTC	<input type="checkbox"/> CTD
<input type="checkbox"/> GS1	<input checked="" type="checkbox"/> RP11	<input type="checkbox"/> RP13		

Homo sapiens CloneFinder Build **37.2**

Assembly GRCh37.p2 - Primary Assembly [Reference Complete][Assembly GCF_000001405.14][Assembly Unit GCF_000001305.13] [Back to search](#)

Chromosome 19 NC_000019.9 45,312,338-45,574,214 bp

Data Summary Download: [Image](#) [Excel](#)

45,312,338 45,350 K 45,400 K 45,450 K 45,500 K 45,574,214

Contigs: NT_011109.16

Components: AC011481.4 AC011489.6

Genes on sequence: BCAM PWR12 CTFP1M RPL10 SFRS16

Transcripts on sequence: NM_001042724.1 NM_002656.2 NM_001294.1 NM_006509.2 NM_007056.2

Ensembl gene annotations: BCAM PWR12 CTFP1M RPL10 SFRS16

Ensembl transcript: RP11-1147010

Feature RP11-1147010
Type Clone
Description(s) Library: RP11
Chrom 19
Chr Pos 45,339,777 - 45,492,883
Contig NT_011109.16
Contig Pos 17,607,995 - 17,761,101
Span 153,107

Clone Library: RP11

Clone Ends
AQ776185.1 span: 407
AQ823421.1 span: 307

Concordant Discordant

Figure 2. The Clone Finder tool. *Top panel:* The Clone Finder homepage with access to clones for a number of genomes in Map Viewer. *Middle panel:* Clone Finder for Homo sapiens Build 37.2 set to find BAC clones from the RP11 library for the region between the genes BCAM and SFRS16. *Bottom panel:* Results shown in the browser with the clone RP11-1147O10 selected showing position information. BAC end sequences are listed and linked to the corresponding Genome Survey Sequence (GSS) records.

Genomes (KEGG) the BioSystems record provides the option to highlight the record in the large pathway diagram at the KEGG site (Figure 3, *Middle panel*). A feature article in the [July 2009 NCBI News](#) describes the BioSystems resource in detail.

New Organisms in UniGene

Five new organisms have builds in UniGene: the hydrozoan (Cnidaria) *Clytia haemisphaerica*; the perigord truffle, *Tuber melanosporum*; the English or Truffle Oak, *Quercus robur*; the two-spotted spider mite, *Tetranychus urticae*; and the salmon louse, *Lepeophtheirus salmonis*.

Clytia haemisphaerica ([Che build information](#), [4,637 clusters](#), [FTP](#)) is a marine hydrozoan in the phylum Cnidaria, the phylum that also contains jellyfish (scyphozoa) and corals (anthozoa). Unlike *Hydra*, the other hydrozoan in UniGene, *Clytia* has a free-swimming medusa stage. Gene and genomic information from *Clytia* has the potential to provide important insights on the evolution of animal body plans.

The perigord truffle ([Tme build information](#), [7,543 clusters](#), [FTP](#)), an ascomycete fungus and the source of the gastronomically highly prized black truffle, and the truffle oak ([Qro build information](#), [7,170 clusters](#), [FTP](#)) are two organisms linked in a symbiotic mycorrhizal association. UniGene sets from these two organisms should support studies of genes involved in the evolution, function, and maintenance of symbiosis.

Two parasitic arthropods of economic importance also join UniGene. The two-spotted spider mite ([Tur build information](#), [7,177 clusters](#), [FTP](#)) is a significant pest of ornamental and horticultural plants. The salmon louse ([Lsl build information](#), [9,363 clusters](#), [FTP](#)) is an ectoparasitic copepod parasite that can cause significant mortality in farmed and wild salmon. These sets may prove helpful in understanding the biology of parasitism and provide targets for control of these pests.

NCBI Databases in Nucleic Acids Research Database Issue

The Nucleic Acids Research 2011 Database Issue contains nine articles about NCBI resources, tools, and databases including Gene, GEO, Epigenomics, CDD and GenBank. Free full-text articles from the database issue are available from PubMed Central and the publisher's site and are linked to the [summaries](#) and [abstracts](#) in PubMed.

dbSNP BLAST Pages Updated

The [dbSNP BLAST](#) page has an updated submission form and output format. The new pages have improved organism selection, chromosome specific database selection, and many of the convenient features of the other BLAST services.

New Mammalian Genomes at NCBI

Updated genome annotations for the rat ([build 4.2](#)), cow ([build 5.2](#)), and a new pig assembly ([build 2.1](#)) are now available for searching and viewing in Entrez, BLAST, the Map Viewer, and for downloading from [genomes area](#) of the FTP site.

Microbial Genomes Update

Sixty-five finished microbial genomes were released during November and December 2010. The original sequence data files submitted to GenBank/EMBL/DDJB are available in the [Bacteria directory](#) in the genomes area of the GenBank FTP site. [RefSeq provisional versions](#) were made for a selected set of 46 these genomes.

Pathways from BioSystems

[Angiotensin receptor Tie2-mediated signaling, organism-specific biosystem](#) (from Pathway Interaction Database)

[Cell surface interactions at the vascular wall, organism-specific biosystem](#) (from REACTOME)

[Class A/1 \(Rhodopsin-like receptors\), organism-specific biosystem](#) (from REACTOME)

[Common Pathway, organism-specific biosystem](#) (from REACTOME)

[Complement and coagulation cascades, organism-specific biosystem](#) (from KEGG)

[Complement and coagulation cascades, conserved biosystem](#) (from KEGG)

[Diabetes pathways, organism-specific biosystem](#) (from REACTOME)

[FOXA2 and FOXA3 transcription factor networks, organism-specific biosystem](#) (from Pathway Interaction Database)

[Formation of Fibrin Clot \(Clotting Cascade\), organism-specific biosystem](#) (from REACTOME)

[Formation of Platelet plug, organism-specific biosystem](#) (from REACTOME)

[G alpha \(q\) signalling events, organism-specific biosystem](#) (from REACTOME)

[GPCR downstream signaling, organism-specific biosystem](#) (from REACTOME)

[GPCR ligand binding, organism-specific biosystem](#) (from REACTOME)

[Gamma-carboxylation of protein precursors, organism-specific biosystem](#) (from REACTOME)

[Gamma-carboxylation, transport, and amino-terminal cleavage of proteins, organism-specific biosystem](#) (from REACTOME)

[Hemostasis, organism-specific biosystem](#) (from REACTOME)

COMPLEMENT AND COAGULATION CASCADES

Genes	Proteins	Small Molecules	Related BioSystems	Citations	Comments
View or save all or selected records in Entrez Gene <input type="button" value="Clear Selections"/> <input checked="" type="button" value="Highlight Selected Records in Source Database"/>					
<input type="checkbox"/>	Gene ID	Gene Symbol	External ID	Name	
<input checked="" type="checkbox"/>	2147	F2	2147	coagulation factor II (thrombin)	
<input type="checkbox"/>	2	A2M	2	alpha-2-macroglobulin	
<input type="checkbox"/>	462	SERPINC1	462	serpin peptidase inhibitor, clade C (antithrombin), member 1	
<input type="checkbox"/>	623	BDKRB1	623	bradykinin receptor B1	
<input type="checkbox"/>	624	BDKRB2	624	bradykinin receptor B2	
<input type="checkbox"/>	629	CFB	629	complement factor B	
<input type="checkbox"/>	710	SERPING1	710	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1	
<input type="checkbox"/>	712	C1QA	712	complement component 1, q subcomponent, A chain	
<input type="checkbox"/>	713	C1QB	713	complement component 1, q subcomponent, B chain	
<input type="checkbox"/>	714	C1QC	714	complement component 1, q subcomponent, C chain	

Page 1 of 7 Displaying Genes 1 - 10 of 69

Figure 3. Linking to BioSystems from the Gene record for F2 coagulation factor II (thrombin) (Gene ID: 2147). *Top panel:* Explicit links to Biosystems from the Gene records. *Bottom panel:* BioSystems record for Complement and coagulation cascades, organism-specific biosystem with Gene ID 2147 selected (bsid83073). Clicking the Highlight link opens the pathway diagram at the KEGG site with F2 highlighted. *Middle panel:* A portion of the KEGG pathway showing the Coagulation cascade.

In addition, 100 microbial whole genome shotgun-sequencing projects were added to GenBank during this period. The original submitted files are available in the [Bacteria_DRAFT](#) directory in the GenBank genomes area. RefSeq provisional versions of 64 of these projects are also available.

All GenBank and RefSeq microbial genomes are incorporated in the NCBI integrated [Entrez](#) search and retrieval system.

New Video on NCBI's YouTube Channel

A [new video](#) that shows how use My NCBI to save searches and set up automated E-mail alerts for new results is now available on [NCBI's YouTube channel](#).

GenBank News

GenBank release 181 is available through the NCBI web and [FTP](#) sites. The current release incorporates data available as of Dec 15, 2010 and contains 122,082,812,719 bases from 129,902,276 sequence records. [Release notes](#) describe the current state of data and upcoming changes.

RefSeq News

RefSeq Release 45 is now available through the Entrez system and can be downloaded from the [FTP site](#). This full release incorporates genomic, transcript, and protein data available as of January 7, 2011 and includes 16,748,646 records from 11,536 different species and strains. The [release notes](#) describe changes since the last release. New in this release is the inclusion of additional features present on the corresponding UniProt/Swiss-Prot record for a subset of RefSeq proteins. These new features are indicated with a Note that identifies the source accession number. An example from [NP_080213.3](#) is shown below.

```
Site          147
              /site_type="phosphorylation"
              /experiment="experimental evidence, no
              additional details recorded"
              /note="Phosphoserine; propagated from
              UniProtKB/Swiss-Prot(Q9D0F4.1)"
```

The [RefSeq Homepage](#) has more information on the RefSeq project.

Journals Database Now a Part of NLM Catalog

The NCBI Journals Database is now part of NCBI [NLM Catalog](#). The NLM Catalog contains the detailed MEDLINE indexing information for the journals in PubMed and other NCBI databases and will maintain the functions of the Journals database.

Announce Lists and RSS Feeds

Eighteen topic-specific mailing lists are available which 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: www.ncbi.nlm.nih.gov/Sitemap/Summary/email_lists.html.

To receive updates on the *NCBI News*, please see: www.ncbi.nlm.nih.gov/About/news/announce_submit.html.

Twelve RSS feeds are now available from NCBI including news on PubMed, PubMed Central, NCBI Bookshelf, LinkOut, HomoloGene, UniGene, and NCBI Announce. Please see: www.ncbi.nlm.nih.gov/feed/.

Users can also stay updated on NCBI's resources on Facebook and Twitter: www.twitter.com/NCBI.

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.