

## GenBank Enters the Megabase Sequence Era

Large-scale sequencing efforts have already produced a number of completely sequenced genomes or chromosomes from a variety of organisms. In making these sequences available, GenBank is charting a course to best serve the varied needs of the scientific community. On the one hand it is scientifically exciting to view the large-scale organization of very long stretches of contiguous DNA. But on the other hand, most biology still focuses on the detailed study of individual genes. Having a database search turn up megabases of DNA surrounding a gene of interest can

often be more of a hindrance than a help.

### New Genome Division

At the annual International Nucleotide Sequence Database Collaborators meeting in April, GenBank, EMBL, and DDBJ agreed on a practical approach to handling megabase sequences. Rather than creating single large entries, genome-size submissions will be divided into several entries, each no more than 350 KB long. These will be assigned to the appropriate GenBank division. "Virtual records" that define the method for assembling the long

sequence will be stored in a new genome division. The individual segments can be assembled by retrieval software so that users can view on demand the complete genome, chromosome, or other unit of interest.

The 350KB limit for any individual database entry is a maximum, not a recommended size, and was selected so as to not break existing molecular biology software tools. Submitters are encouraged to submit entries below this limit, corresponding to the "natural" units in which the sequencing is done, often cosmid size pieces, or entries containing only a few genes.

## New Entrez Release Takes Graphical View

The October release of Network Entrez and Entrez on CD-ROM will include a graphical viewer for displaying the locations of features annotated on sequences. A graphical overview is helpful for understanding large or complicated sequence entries and is much easier to interpret than a list of numerical positions shown in text report formats. The capability is essential for viewing sequences of entire chromosomes and genomes, and associated maps, available through the new genome division in Network Entrez. Although access to the genome division is not possible on the CD-ROM, all other graphical viewing functions are there.

A tabbed folder approach to selecting alternate report formats has also been added, making it very easy to move quickly between text and graphical display formats.

### 3D Structure

For Network Entrez users, the new release also includes an explicit 3D structure database derived from crystallographic and NMR data in PDB, the Brookhaven Protein Databank. As with the sequence and bibliographic databases, the structure database may be queried directly, using specific fields such as author names or text terms, to check for structure data on a specific protein or nucleic acid. Structure data may

The sequence database collaboration has already defined the virtual record format, which contains feature table information on how to assemble the individual entries into a single contiguous sequence representing

*Continued on page 2*

### IN THIS ISSUE

GenBank Enters Megabase Era .....	1
Entrez Takes Graphical View .....	1
GenBank Taxonomy .....	2
BankIt Submissions Mount .....	3
NCBI Data by FTP .....	3
Recent Publications .....	4
Frequently Asked Questions .....	5
GenBank Services .....	7

*Continued on page 4*

NCBI News is distributed three times a year. We welcome communication from users of NCBI databases and software and invite suggestions for articles in future issues. Send correspondence and suggestions to *NCBI News* at the address below.

NCBI News  
National Library of Medicine  
Bldg. 38A, Room 8N-803  
8600 Rockville Pike  
Bethesda, MD 20894  
Phone: (301) 496-2475  
Fax: (301) 480-9241  
E-mail: [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)

*Editors*  
Dennis Benson  
Barbara Rapp

*Design Consultant*  
Troy M. Hill

*Photography*  
Karlton Jackson

*Editing, Graphics, and Production*  
Veronica Johnson  
Wendy B. Osborne

In 1988, Congress established the National Center for Biotechnology Information as part of the National Library of Medicine; its charge is to create automated systems for storing molecular biology, biochemistry, and genetics data, and to perform research in computational molecular biology.

The contents of this newsletter may be reprinted without permission. The mention of trade names, commercial products, or organizations does not imply endorsement by NCBI, NIH, or the U.S. Government.

NIH Publication No. 95-3272

ISSN 1060-8788

## *GenBank Enters the Megabase Sequence Era, continued from page 1*

the complete megabase sequence. The sequence databases will individually be experimenting with ways to present these sequences to users this year, then share experiences at next year's meeting and determine the optimal path for future development.

### **Complete Sequence**

GenBank will be offering the megabase sequences in a number of forms on our FTP site and through our search services.

The virtual records will be available in GenBank flat-file format as well as ASN.1. These records are actually quite small because they contain no sequence, only information about how to put other records together to make the megabase sequence. The NCBI will also instantiate the virtual records by filling in the sequence and feature tables according to the assembly instructions, creating a single huge entry. These large composite entries will be available

in FASTA format and GenBank flat-file format on NCBI's FTP site ([ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)) in the *genbank/genomes* directory. The complete *Haemophilus influenzae* sequence is currently available in this directory.

### **Graphical Views**

Network Entrez users will have the added functionality of a graphical sequence viewer. This new feature will present graphical views of sequences, and any associated genetic and physical maps, on demand. One will be able to view a schematic of the whole megabase sequence, and then look in detail at a subregion. Any subregion can then be selected for a more detailed graphical view of biological features annotated on that region.

These and other new views and services will become available to the public over the next year. We encourage your comments and suggestions during this period. ■



## **GenBank Taxonomy: Is a Rabbit a Fish?**

Users are starting to notice the taxonomy changes in GenBank. This is the result of a project started 2 years ago to build a uniform, phylogenetically based taxonomy for GenBank. Because the taxonomy is based on phylogenetics, some of the relationships appear unusual at first glance, such as the inclusion of humans and rabbits under bony fish.

In the current classification, for example, the Gnathostomata (jawed vertebrates) include all vertebrates except the lampreys and hagfish; the Osteichthyes (bony vertebrates) include all Gnathostomata except

the cartilaginous fish; and the Sarcopterygii (lobe-finned fishes) include all Osteichthyes except the ray-finned fishes.

### **Phylogenetic Approach**

The impetus for the project was the need for a consistent and comprehensive sequence-based taxonomy to process and query the sequence databases built at NCBI. With the support of the international collaborating databases, EMBL and DDBJ, the GenBank taxonomy was developed by merging and unifying taxonomic data from a variety of sources. The project is not intended

*Continued on page 6*

# BankIt Submissions Mount

**B**ankIt, the World Wide Web (WWW) tool for submitting sequences to GenBank, has been used to submit more than 7,000 GenBank entries and now accounts for more than two-thirds of new submissions each month. Since its introduction this past February, a number of improvements have been made to meet the needs expressed by our users.

## More Than 30,000 Bases

BankIt now accepts sequences longer than 30,000 nucleotides. Although most, if not all, WWW browsers still have an inherent limitation of approximately 30,000 characters per input window, BankIt circumvents this by first asking how many nucleotides you intend to submit. The appropriate number of DNA sequence input windows, each with a 30KB capacity, is then incorporated into the BankIt submission form.

Note that there is still an upper limit of 350,000 nucleotides for individual GenBank records, as agreed by the international collaboration of DNA sequence databases (see article on page 1). Sequences larger than 350,000 should be broken down into smaller segments and submitted as separate entries that will be linked together by software.

## Your BankIt ID Number

If your Web client crashes or you forgot to save a copy of your submission file, all is not lost. NCBI maintains a BankIt transaction log and assigns an identification number to each BankIt submission. So if you ever need to retrieve an incomplete submission, just tell us your BankIt ID number. We will e-mail the submission to you in HTML format, then you can reload it into BankIt and complete it. Note that your BankIt ID is *not* your GenBank accession number.

## Updating Existing Entries

BankIt can now be used to modify or update any of your own GenBank records, regardless of whether BankIt was used to submit them originally. Choose Update from the BankIt opening screen, then enter your accession number. If the record is in the public release of GenBank, BankIt will display it. If your record is being held confidential, it will not be displayed, but you can still specify the modifications to be made. If you wish to make modifications to a very recent submission for which you do not yet have an accession number, you can use your BankIt ID number instead.

## Saving a BankIt File

A new button called Save This Form has been added at the bottom of the BankIt submission form. Saving a copy of your submission is useful if you have several similar sequences to submit, or if you want to save an incomplete submission and come back to it later. When you have completed each BankIt submission, we recommend that you save a copy in HTML format for your records.

*Continued on page 8*

## NCBI Data by FTP

The NCBI FTP site contains a variety of directories with publicly available databases and software. The available directories include "repository", "genbank", "entrez", "toolbox", and "pub".

The **repository** directory makes a number of molecular biology databases available to the scientific community. This directory includes databases such as PIR, Swiss-Prot, CarbBank, AceDB, and FlyBase.

The **genbank** directory contains files with the latest full release of Genbank, the daily cumulative updates, and the latest release notes.

The **entrez** directory contains the Entrez executable programs for accessing CD-ROM data on a variety of platforms. It also contains client software for Network Entrez.

The **toolbox** directory contains a set of software and data exchange specifications that are used by NCBI to produce portable software, and includes ASN.1 tools and specifications for molecular sequence data.

The **pub** directory offers public-domain software, such as BLAST (sequence similarity search program), MACAW (multiple sequence alignment program), and Authorin submission software for Mac and PC systems. Client software for Network BLAST is also included in this directory.

Data in these directories can be transferred through the Internet by using the Anonymous FTP program. To connect, type: **ftp ncbi.nlm.nih.gov** or **ftp 130.14.25.1**. Enter **anonymous** for the login name, and enter your e-mail address as the password. Then change to the appropriate directory. For example, change to the repository directory (cd repository) to download specialized databases.



---

---

## Selected Recent Publications by NCBI Staff

**Baxevanis, AD, SH Bryant, and D Landsman.** Homology model building of the HMG-1 box structural domain. *Nucleic Acids Res* 23(6):1019-29, 1995.

**Baxevanis, AD, and D Landsman.** The HMG-1 box protein family: classification and functional relationships. *Nucleic Acids Res* 23(9):1604-13, 1995.

**Boguski, MS.** Molecular medicine: hunting for genes in computer data bases. *N Engl J Med* 333(10):645-47, 1995.

**Boguski, MS, and GD Schuler.** ESTablishing a human transcript map. *Nat Genet* 10:369-71, 1995.

**Bryant, SH, and SF Altschul.** Statistics of sequence-structure threading. *Curr Opin Struct Biol* 5:236-44, 1995.

Bussey, H, DB Kaback, W Zhong, DT Vo, MW Clark, N Fortin, **BFF Ouellette**, R Keng, AB Barton, Y Su, CK Davies, and RK Storms. The sequence of chromosome I from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 92:3809-13, 1995.

Castonguay, LA, **SH Bryant**, PW Snow, and JS Fetrow. A proposed structural model of domain 1 of faciclin II neural cell adhesion protein based on an inverse folding algorithm. *Protein Sci* 4:472-83, 1995.

Klaassen, VA, M Boeshore, **EV Koonin**, T Tian, and BW Falk. Genome structure and phylogenetic analysis of lettuce infectious yellows virus, a whitefly-transmitted, bipartite closterovirus. *Virology* 208:99-110, 1995.

**Landsman, D,** and AP Wolffe. Common sequence and structural features in the heat shock factor and ets families of DNA-binding domains. *Trends Biochem Sci* 20(6):225-6, 1995.

Rouviere PE, A De Las Penas, J Mescas, CZ Lu, **KE Rudd**, and CA Gross. rpoE, the gene encoding the second heat-shock sigma factor, sigmaE, in *Escherichia coli*. *EMBO J* 14:1032-42, 1995.

Sanderson KE, A Hessel, and **KE Rudd.** Genetic map of *Salmonella typhimurium*, edition VIII. *Microbiol Rev* 59:241-303, 1995.



### *Entrez Takes Graphical View, continued from page 1*

then be viewed in 3D, with real-time rotation, using the public domain graphics programs RasMol or Kinemage. Entrez itself simply writes structure documents in the format required by these programs. Future versions, however, will invoke an integrated 3D structure viewer directly from Network Entrez. The graphical interface is not yet available in WebEntrez.

### Daily Updates

Network Entrez and WebEntrez are now updated daily with newly released GenBank, EMBL, DDBJ, and GSDB records. New entries can be retrieved by searching on any of the Entrez data fields. Sequence neighbors, however, lag slightly behind the availability of the new records themselves, due to the extensive processing required. The sequence neighbors are currently updated weekly. Protein sequence entries from Swiss-Prot, PIR, PDB, and PRF are added whenever NCBI obtains their public releases. The MEDLINE subset is updated weekly.

### No More Registration

It is no longer necessary to register your computer's IP address prior to using Network Entrez. However, users at sites that are already registered will still see the name of their local administrator when they connect. For assistance with network access problems, please continue to consult first with your local systems support staff. Contact NCBI for bug reports or assistance with using the features of Network Entrez.

### Links to JBC Online

WebEntrez now contains links to and from JBC Online, the on-line version of the *Journal of Biological*

*Chemistry*, beginning with the April 14, 1995, issue. Starting from WebEntrez, select the MEDLINE data set, then locate a record that was published in JBC. Click on the JBC button to link to JBC Online and see the full text of the article.

Starting from JBC Online, links to GenBank are available in articles that report a new sequence. When a linked accession number appears in an article, the GenBank link is highlighted. Click on the link to connect to Entrez and see the full GenBank record. Links are also available from many of the references in JBC Online articles. Click on any reference that includes a MEDLINE link to connect to Entrez and see the MEDLINE abstract.

As other electronic journals become available on-line, NCBI intends to make similar links. JBC Online is available from The Highwire Press through its WWW site (<http://highwire.stanford.edu/jbc>). This service is still in the development stage, and access is free of charge for a trial period.

### CD-ROM Expands to Five Discs

The December 1995 release of Entrez on CD-ROM will require five discs. Due to the influx of EST data, the sequence databases are growing at a faster rate than was anticipated a year ago. A price increase will accompany each expansion to an additional disc.

The NCBI encourages subscribers to switch to either Network Entrez or WebEntrez. For more information on Internet versions of Entrez, contact [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov) ■



---

## Frequently Asked Questions

---

*I recently used BankIt to submit a sequence to GenBank, but I haven't received any confirmation. My BankIt number was 12345. Should I do anything else?*

Contact GenBank User Support, and they will check the BankIt transaction log to confirm that a completed submission was received. Incomplete submissions can be retrieved and e-mailed to you for completion. On the BankIt Revision Page, note that if you do not explicitly click on the Submit to GenBank option before pressing the BankIt button for the last time, your submission remains incomplete.

*A previously unmapped EST maps to the region I'm working in, assuming no duplications. Should I send this information to dbEST?*

Yes, dbEST does accept mapping data for EST sequences submitted by someone else. Basically, you will submit four small files, one for your contact information, one for the mapping method, one for a citation to the mapping method, and one for the map data itself. NCBI does have special formatting requirements for EST data, so obtain detailed instructions from [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)

*When using the RETRIEVE e-mail server, how can I get around the 1,000- and 50,000-line limits on my output?*

The MAXLINES command allows you to change the default limit of 1,000 lines to any number up to the maximum of 50,000. The STARTDOC command allows you to obtain output in several batches. This can be used to circumvent the 50,000-line maximum as well as any limits in the size of mail messages you are able to receive at your end.

*With the BLAST e-mail server, is there any way to be sure you got my search and to find out how long the queue is?*

The new ACKNOWLEDGE command allows you to receive a notice after any length of time you specify. If your search is not completed within that time period, the BLAST server will send an e-mail message informing you of your position in the processing queue. See the BLAST documentation for more detail.

*How can I find out the total number of entries and nucleotides in GenBank?*

These numbers are published at the beginning of the GenBank Release Notes prepared for each release and are available from NCBI's Anonymous FTP site ([ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)) in the directory "genbank". The name of the file is `gbrel.txt`.

*Is the full-sequence of Haemophilus influenzae available?*

Yes, the full contiguous sequence is available on the NCBI FTP site in the `genbank/genomes` directory. The sequence is presented in both FASTA format (1.8MB) and GenBank flat-file format (3.8MB) and as compressed and uncompressed files.

*Authorin doesn't work with my new Mac. What's wrong?*

Authorin only runs with 24-bit addressing. If your Mac allows you to select 24-bit or 32-bit mode, select 24-bit mode, then restart before using Authorin. If you have a newer Mac that only uses 32-bit addressing, you'll need to switch to BankIt on the Word Wide Web. ■

to produce an international standard or official classification, but specifically to support the sequence databases.

Another important factor was the importance of taxonomic relationships to sequence similarities. Because a strictly phylogenetic approach more closely reflects evolutionary history than does classical taxonomy, it is well suited for applications associated with sequence databases. Users, for example, are interested in determining the level of specificity of a particular probe or in identifying a distantly related organism that has the same gene that they have isolated.

Entrez now includes a taxonomy search mode, which can be used to explore the GenBank classification and do tree-based retrieval of sequence data. Users can retrieve sequences based on scientific name and hierarchical classification, then browse upward and downward through the phylogenetic tree to retrieve sequences from related taxa.

### 15,500 Species in GenBank

The taxonomic relationships are based on publications wherever possible, and literature references are provided so that users will be able to independently assess the logic of the GenBank classification. In order to be comprehensive, all organisms in GenBank must have an entry in the tree. In this regard, the taxonomy is driven by the organisms being sequenced rather than all organisms that exist. As of August, there were 1,920 family nodes, 5,588 genus nodes, 15,511 species nodes, and 1,903 nodes below the species level represented in GenBank. An average of 10 new organisms are added each day.

Three NCBI scientists experienced in taxonomy and molecular biology—Scott Federhen, Andrzej Elzanowski, and Detlef Leipe—maintain the taxonomy internally. Additionally, outside molecular

biologists and taxonomists serve as curators and provide expert review and consultation (see box). The list of advisors will continue to grow over the next few months. ■

#### Contributors to Taxonomy Project

Michael Ashburner, European Bioinformatics Institute: *dipterans*  
Gerhard Baechli, University of Zurich: *dipterans*  
James G. Baldwin, University of California at Riverside: *nematodes*  
Meredith Blackwell, Louisiana State University: *fungi*  
Bruce Campbell, Agricultural Research Service, USDA: *true bugs*  
Russell Chapman, Louisiana State University: *green algae*  
Douglas Eernisse, California State University: *metazoa*  
Mark Farmer, University of Georgia: *euglenoids, kinetoplastids, and trichomonads*  
Kristian Fauchald, Smithsonian Institution: *polychaetes*  
Suzanne Fredericq, Smithsonian Institution: *red algae*  
Wilson Freshwater, University of Miami: *red algae*  
Walter Gams, Centraalbureau voor Schimmelcultures (The Netherlands): *fungi*  
Gerald J. Gastony, Indiana University: *ferns*  
William J. Hahn, Smithsonian Institution: *flowering plants*  
William C. Hart, Jr., Smithsonian Institution: *decapod crustaceans*  
David Hillis, University of Texas: *chordates*  
Eugene Koonin, NCBI: *viruses*  
Phil Lambert, Royal British Columbia Museum: *sea cucumbers*  
Jon L. Norenburg, Smithsonian Institution: *ribbon worms*  
Richard Olmstead, University of Colorado: *dicotyledons*  
Gary Olsen, University of Illinois at Urbana-Champaign: *bacteria*  
David Patterson, University of Sydney: *stramenopiles*  
Norman Pieniazek, Centers for Disease Control and Prevention: *microsporidians*  
Norman Platnick, American Museum of Natural History: *spiders*  
Jerry Powell, University of California at Berkeley: *moths*  
Harry M. Savage, Centers for Disease Control and Prevention: *mosquitos*  
Jeffrey Jon Shaw, Belem Research Project (Brazil): *leishmanias*  
David Sissom, West Texas A&M University: *scorpions*  
Alan R. Smith, University of California at Berkeley: *ferns*  
Mitchell Sogin, Marine Biological Laboratory at Woods Hole: *protists*  
Felix Sperling, University of California at Berkeley: *butterflies*  
John Taylor, University of California at Berkeley: *fungi*  
Robert Van Syoc, California Academy of Sciences: *cirripeds*  
Steven J. Wagstaff, Landcare Research New Zealand Ltd.: *dicotyledons*  
George R. Zug, Smithsonian Institution: *reptiles*

# GenBank: Easy Deposits, Unlimited Withdrawals, High Interest

It's easy—and free—to contribute sequences to GenBank and search the database. This table summarizes the data submission and search services available from NCBI.

Service	Purpose	How To Use/How To Get Help
<b>Information</b>	Obtain general information about NCBI databases and services.	Send e-mail to <a href="mailto:info@ncbi.nlm.nih.gov">info@ncbi.nlm.nih.gov</a> or call GenBank User Services at (301) 496-2475.
<b>GenBank submissions</b>	Submit new sequences to GenBank.	For information or technical assistance: <a href="mailto:info@ncbi.nlm.nih.gov">info@ncbi.nlm.nih.gov</a>
Authorin software	Prepare new or updated GenBank entry.	Send a new submission by e-mail: <a href="mailto:gb-sub@ncbi.nlm.nih.gov">gb-sub@ncbi.nlm.nih.gov</a>  To obtain software for Mac or PC, send request to: <a href="mailto:authorin@ncbi.nlm.nih.gov">authorin@ncbi.nlm.nih.gov</a>
BankIt on WWW	Prepare and submit new GenBank entry over the Internet, using the World Wide Web.	For information on compatible WWW browsers: <a href="mailto:info@ncbi.nlm.nih.gov">info@ncbi.nlm.nih.gov</a>  To access BankIt through NCBI Home Page: <a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
<b>GenBank updates</b>	Correct or update an existing sequence; request release of published data.	Send an update request by e-mail: <a href="mailto:update@ncbi.nlm.nih.gov">update@ncbi.nlm.nih.gov</a>
<b>E-mail servers</b>		
<a href="mailto:retrieve@ncbi.nlm.nih.gov">retrieve@ncbi.nlm.nih.gov</a>	Retrieve GenBank and other sequence database records from an e-mail server based on any text term, including accession number, author name, locus, gene name, etc.	To receive documentation, send a message containing only the word HELP to the server address. For personal assistance, send questions to: <a href="mailto:retrieve-help@ncbi.nlm.nih.gov">retrieve-help@ncbi.nlm.nih.gov</a>
<a href="mailto:blast@ncbi.nlm.nih.gov">blast@ncbi.nlm.nih.gov</a>	Perform a sequence similarity search of GenBank and other sequence databases using the BLAST algorithm.	To receive documentation, send a message containing only the word HELP to the server address. For personal assistance, send questions to: <a href="mailto:blast-help@ncbi.nlm.nih.gov">blast-help@ncbi.nlm.nih.gov</a>
<b>Internet applications</b>	“Client-server” programs, in which client program on local PC, Mac, or Unix workstation queries NCBI server via the network.	All NCBI network applications require Internet access and locally installed TCP/IP software.
Network Entrez	Point-and-click retrieval system for PCs running Windows, Macs, and Unix workstations. Provides text-based searching of sequence databases and a sequence-related subset of MEDLINE.	To obtain client software, send e-mail to: <a href="mailto:net-info@ncbi.nlm.nih.gov">net-info@ncbi.nlm.nih.gov</a>
Network BLAST	BLAST client for similarity searching for PC (DOS), Mac, Unix, and VMS workstations.	To register and obtain client software, send e-mail to: <a href="mailto:blast-help@ncbi.nlm.nih.gov">blast-help@ncbi.nlm.nih.gov</a>
World Wide Web access	WWW access to NCBI databases and search services, including BankIt for GenBank submissions and Web versions of RETRIEVE, BLAST, and Entrez.	For information on compatible WWW browsers: <a href="mailto:info@ncbi.nlm.nih.gov">info@ncbi.nlm.nih.gov</a>  To access NCBI Home Page: <a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
<b>Anonymous FTP:</b> <a href="http://ncbi.nlm.nih.gov">ncbi.nlm.nih.gov</a>	Obtain GenBank releases, NCBI software, and various molecular biology databases.	Login as “anonymous” (unquoted) and enter your e-mail address as your password.
<b>CD-ROMs</b>	For users who do not have Internet access or who prefer a local copy of databases.	For information about subscriptions, send e-mail to: <a href="mailto:info@ncbi.nlm.nih.gov">info@ncbi.nlm.nih.gov</a>
Entrez (GPO list ID: ENT)	CD-ROM version of Network Entrez. Annual subscription (6 issues per year).	For technical assistance, send e-mail questions to: <a href="mailto:entrez@ncbi.nlm.nih.gov">entrez@ncbi.nlm.nih.gov</a>
GenBank (GPO list ID: NCBIF)	GenBank in “flat-file” format, as used by some commercial and academic software. Annual subscription (6 issues per year).	Send e-mail to: <a href="mailto:info@ncbi.nlm.nih.gov">info@ncbi.nlm.nih.gov</a>

To save a submission form, press the Save This Form button, then click on BankIt. Netscape and MacWeb users will be prompted by their browser to enter a file name, then the file will be saved automatically in HTML format on their local system. MacWeb users need to include .html as the filename extension. Saving is not completely automatic for Unix-based Mosaic users. They need to press Save This Form, then use Mosaic's "Save As" feature to name and save the file in HTML format. When the saving is completed, all users should turn the Save This Form button off by clicking on it again, then pressing the BankIt button to continue.

### **Submission Tips**

#### ***Change Status to Submit!***

The most important tip is to change the status of your submission from Modify to Submit before clicking the BankIt button for the last time. Otherwise, your submission is still incomplete. You will know that your submission is complete when the BankIt window displays a thank you message. You will also receive an e-mail acknowledgment thanking you for your submission.

#### ***Don't Forget the Annotations!***

The initial BankIt form does not provide for biological annotations.

However, once you enter the initial information and click the BankIt button, you will have an opportunity to review your entry and specify the number of coding regions, structural RNA features, or other biological features you wish to add. Click the BankIt button again, and you'll be able to enter the additional information at the end of your original BankIt submission form.

### **Help**

If you have any questions on using BankIt, contact GenBank User Services at [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov) or at (301) 496-2475. ■

---

DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Public Health Service, National Institutes of Health  
National Library of Medicine  
National Center for Biotechnology Information  
Bldg. 38A, Room 8N-803  
8600 Rockville Pike  
Bethesda, Maryland 20894

---

Official Business  
Penalty for Private Use \$300

FIRST-CLASS MAIL  
POSTAGE & FEES PAID  
PHS/NIH/NLM  
BETHESDA, MD  
PERMIT NO. G-763