

# **NUCLEOTIDE SEQUENCES 1986/1987**

**VOLUME VIII**  
**DATABASE DIRECTORY  
AND MASTER INDICES**

**A compilation from the  
GenBank<sup>®</sup>  
and  
EMBL  
data libraries**

0042342

# NUCLEOTIDE SEQUENCES 1986/1987

1551CHA

## VOLUME VIII DATABASE DIRECTORY AND MASTER INDICES

A Compilation from the  
**GenBank®**  
and  
**EMBL**  
data libraries

Compiled by

Edwin J. Atencio,\* Howard S. Bilofsky,† June Bossinger,† Christian Burks,\* Graham N. Cameron,† Michael J. Cinkosky,\* Carol E. England,\* Victor I. Esekogwu,\* James W. Fickett,\* Brian T. Foley,\* Walter B. Goad,\* Gregory H. Hamm,‡ David J. Hazledine,‡ Patricia Kahn,‡ Leslie Kay,\* Frances I. Lewitter,† Natalie Lopez,\* Kersti A. MacInnes,\* Mia J. McLeod,\* Deborah L. Melone,† Gerald Myers,\* Debra Nelson,\* Judith L. Nial,‡ Joanna K. Norman,\* Eric D. Rasmussen,\* Andrea A. Revels,\* Wayne P. Rindone,† Carol R. Schermer,\* Maura T. Smith,\* Guenter Stoesser,‡ C. David Swindell,† Brian L. Trujillo,\* and Chang-Shung Tung\*

\* *GenBank*  
T-10 Mail Stop K710  
Los Alamos National Laboratory (LANL)  
Los Alamos, New Mexico 87545

† *GenBank*  
BBN Laboratories Incorporated  
10 Moulton Street  
Cambridge, Massachusetts 02238

‡ *EMBL Nucleotide Sequence Data Library*  
European Molecular Biology Laboratory  
Postfach 10 22 09  
D-6900 Heidelberg  
Federal Republic of Germany



ACADEMIC PRESS, INC.  
Harcourt Brace Jovanovich, Publishers  
Orlando San Diego New York Austin  
Boston London Sydney Tokyo Toronto

COPYRIGHT © 1987 BY ACADEMIC PRESS, INC.  
ALL RIGHTS RESERVED.  
NO PART OF THIS PUBLICATION MAY BE REPRODUCED OR  
TRANSMITTED IN ANY FORM OR BY ANY MEANS, ELECTRONIC  
OR MECHANICAL, INCLUDING PHOTOCOPY, RECORDING, OR  
ANY INFORMATION STORAGE AND RETRIEVAL SYSTEM, WITHOUT  
PERMISSION IN WRITING FROM THE PUBLISHER.

ACADEMIC PRESS, INC.  
Orlando, Florida 32887

United Kingdom Edition published by  
ACADEMIC PRESS INC. (LONDON) LTD.  
24-28 Oval Road, London NW1 7DX

By purchasing or otherwise obtaining *Nucleotide Sequences 1986/1987*, recipient understands that the information contained in this compendium, which has been produced from the information contained in the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Data Library and the GenBank® database ("the information"), has come from a variety of sources, published and perhaps unpublished. The information has been deposited in the GenBank® database and the EMBL Nucleotide Sequence Data Library, and it has been reproduced for inclusion in this compendium via a reliable and quality controlled procedure, but no such process is infallible. Therefore, Academic Press, Inc. (AP), Bolt Beranek and Newman Inc. (BBN), Los Alamos National Laboratory (LANL), The European Molecular Biology Laboratory (EMBL), and the U.S. Government make no representations or warranties regarding the content or accuracy of the information. By way of example, but not of limitation, AP, BBN, LANL, EMBL, and the U.S. Government make no representation or warranties of merchantability or fitness for a particular purpose, or that the use of the information will not infringe any patent, copyright, trade secret, or trademark of any third person. AP, BBN, LANL, EMBL, and the U.S. Government accept no responsibility for any expenses, losses, or action incurred or undertaken by the recipient as a result of the receipt or use of the information.

Note that GenBank® is a registered trademark for the Genetic Sequence Data Bank established by BBN and LANL under contract with the U.S. National Institutes of Health and should be used only in that context.

Information from this compendium may be duplicated, reproduced, or otherwise used by the recipient, but in no event may the GenBank® trademark be associated with such re-generated information, and in no event shall there be any remedy furnished by AP, BBN, LANL, EMBL, or the U.S. Government for such re-generated information, including but not limited to financial remuneration or technical interaction.

Please note that the proper attribution of *Nucleotide Sequences 1986/1987* as the source of your data and the public availability of this information in computer-readable form from BBN and EMBL will be appreciated.

#### Library of Congress Cataloging in Publication Data

Nucleotide sequences 1986/1987.

Includes indexes.

Contents: v. 1. Primates — v. 2. Rodents — v. 3.

Other vertebrates and invertebrates — [etc.]

1. Nucleotide sequence—Tables—Collected works.

I. Atencio, Edwin J. II. GenBank. III. Europæiske molekylærbiologiske laboratorium. IV. BBN

Laboratories. V. Los Alamos National Laboratory.

QP 625.N89N85 1987 547.7'9 87-1782

ISBN 0-12-512518-6 (v. 8 : alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA

0042342

**NUCLEOTIDE  
SEQUENCES  
1986/1987**

**VOLUME VIII  
DATABASE DIRECTORY AND MASTER INDICES**

**Academic Press Rapid Manuscript Reproduction**

# Contents

---

## VOLUME VIII DATABASE DIRECTORY AND MASTER INDICES

Preface	vii
Introduction	ix
Database Directory	DIRECTORY-1 to DIRECTORY-115
Technical Appendix A: Entry Name and Molecule Type Conventions	APPENDIX-1 to APPENDIX-2
Technical Appendix B: Reference Citation Conventions	APPENDIX-3 to APPENDIX-4
Technical Appendix C: Sites and Features Tables	APPENDIX-4 to APPENDIX-6
Technical Appendix D: Sequence Representation Conventions	APPENDIX-7
Technical Appendix E: EMBL and GenBank Format Comparison	APPENDIX-8 to APPENDIX-12
Master Keyword Phrase Index	KEYWORD-1 to KEYWORD-31
Master Taxonomic Classification Index	TAXONOMY-1 to TAXONOMY-34
Master Journal Citation Index	JOURNAL-1 to JOURNAL-44
Master Author Index	AUTHOR-1 to AUTHOR-87
Master Accession Number Index	ACCESSION-1 to ACCESSION-28
Master EMBL Entry Index	EMBL-1 to EMBL-21
Master GenBank Entry Index	GENBANK-1 to GENBANK-28

## SEQUENCE ENTRIES FOUND IN VOLUMES I, II, III, IV, V, VI, and VII

### VOLUME I. PRIMATES

#### Section 1. Primate Sequences

### VOLUME II. RODENTS

#### Section 2. Rodent Sequences

### VOLUME III. OTHER VERTEBRATES AND INVERTEBRATES

#### Section 3. Other Mammalian Sequences

#### Section 4. Other Vertebrate Sequences

#### Section 5. Invertebrate Sequences

### VOLUME IV. PLANTS AND ORGANELLES

#### Section 6. Plant Sequences

#### Section 7. Organelle Sequences

### VOLUME V. BACTERIA AND BACTERIOPHAGE

#### Section 8. Bacterial Sequences

#### Section 9. Bacteriophage Sequences

### VOLUME VI. VIRUSES

#### Section 10. Viral Sequences

**VOLUME VII. STRUCTURAL RNA, SYNTHETIC, AND UNANNOTATED SEQUENCES**

**Section 11. Structural RNA Sequences**

**Section 12. Synthetic Sequences**

**Section 13. Unannotated Sequences**

# Preface

---

This eight-volume compendium of nucleotide sequences found in the GenBank and EMBL databases is the third edition resulting from the combined efforts of all of the technical and administrative staff at Los Alamos National Laboratory, the European Molecular Biology Laboratory, and BBN Laboratories Incorporated listed on the title page. Both the EMBL and GenBank databases have continued to grow at a remarkable rate, with each database doubling in size nearly once each year. We have organized this compendium in eight self-contained volumes, each of which is available separately. The first seven volumes each contain the same introductory and explanatory material, one or more sections of sequence entries, and several indices to the entries in that volume. Volume VIII contains a database directory and master indices to all of the volumes.

As a result of comments and suggestions we received in response to the previous edition, we have made several improvements in this edition. We hope that some slight adjustments in the layout and presentation of the sequence entries, including increasing use of mixed-case text and improvements in punctuation, will result in making them more easily readable than in the past.

Both databases are available in a variety of computer-readable forms. Additional information about obtaining the GenBank database can be obtained by writing to

*Genbank*  
*BBN Laboratories Incorporated*  
*10 Moulton Street*  
*Cambridge, Massachusetts 02238*  
*USA*

Further information about the EMBL Nucleotide Sequence Data Library can be obtained by writing to

*EMBL Nucleotide Sequence Data Library*  
*European Molecular Biology Laboratory*  
*Postfach 10 22 09*  
*D-6900 Heidelberg*  
*Federal Republic of Germany*

Wayne P. Rindone  
Cambridge, Massachusetts  
November 17, 1986



# Introduction

## Outline

1. Introduction
  - 1.1 Description of the compendium
  - 1.2 The two databases
  - 1.3 New features of this edition
2. Contents of the Compendium
  - 2.1 General organization of the compendium
  - 2.2 Finding an entry
3. How to Read an Entry
  - 3.1 Summary of the entry fields
  - 3.2 The fields in detail
4. Two Sample Entries

## 1. Introduction

Nucleotide Sequences 1986/1987 is the third database compendium published as one result of a unique international collaboration between two leading nucleotide sequence data libraries, one based in the United States and one in Europe. The two databases are the EMBL Nucleotide Sequence Data Library, established by the European Molecular Biology Laboratory (EMBL), and the GenBank(R) Genetic Sequence Data Bank, which is a U.S. Government-sponsored nucleic acid sequence repository. Both databases serve molecular biologists and other investigators worldwide by collecting the large number of reported DNA and RNA sequences and making them available in computer-readable form. The primary distribution medium for both databases is magnetic tape.

The data in the compendium reflect the information found in GenBank Release 44.0 of August 1986. This information has been combined with the data included in EMBL Release 8.0, which was made available in May 1986. Regularly updated distribution tapes containing the EMBL Sequence Data Library are available four times annually. A new set of distribution tapes containing the entire GenBank database is also made available four times annually, and update tapes containing only entries that have been added or changed are available midway between each full GenBank release.

The sequences in this compendium are also available from GenBank on floppy diskettes. Because of limited storage capacity, only the sequences, some basic identifying information, and some of the biological annotations are included on this distribution medium. The remaining annotated information can be found in the compendium.

The GenBank database is available online on the DRR/NIH/PROPHET computer system, which can be accessed over Telenet, an international telecommunications network. The online database is updated every six weeks on the same schedule as the magnetic tape releases. This online service also provides users with access to the GenBank Software Clearinghouse, which contains information about commercially available software packages for analyzing and manipulating sequences.

For more information on the services provided by the GenBank and EMBL sequence libraries, please write:

GenBank  
BBN Laboratories Inc.  
10 Moulton St.  
Cambridge, MA 02238  
USA

European Molecular Biology Laboratory  
Nucleotide Sequence Data Library  
Postfach 10.2209  
D-6900 Heidelberg  
West Germany

## 1.1 Description of the compendium

The printed compendium makes the entire collection of information in both databases available to every member of the scientific community who wishes to use it, including investigators without access to computers. This compendium, drawn from the American and European databases, is the third printed compilation of substantially all nucleic acid sequences reported since 1967. These sequences and their associated annotations have been compiled from the published literature and from direct submissions from the authors by the GenBank staff at Los Alamos National Laboratory and by the EMBL data library staff at EMBL.

Although the format chosen for entries in the printed compendium differs somewhat from that in either database, every entry contains information contributed both by EMBL and by GenBank. The final preparation of the data in the compendium was carried out by the GenBank staff at BBN Laboratories Incorporated (BBN); therefore, the format and conventions used in the compendium are somewhat closer to those used in the GenBank database than to those used in the EMBL data library. Technical Appendix E illustrates how the compendium format relates to the formats used in the two databases from which it was constructed. One of the goals of the collaboration between GenBank and EMBL is continued movement toward common standards and conventions for the two databases.

## 1.2 The two databases

The EMBL Nucleotide Sequence Data Library was established in 1980 by the European Molecular Biology Laboratory, an international center of fundamental research with its main emphasis in the fields of cell biology, molecular structures, differentiation, and instrumentation. EMBL, whose headquarters is in Heidelberg, Germany, is currently funded by the following member states: Austria, Denmark, France, Federal Republic of Germany, Finland, Greece, Israel, Italy, the Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom. The first release of the EMBL data library was in April 1982.

The GenBank database was created in 1982 by the National Institute of General Medical Sciences (NIGMS) of the U.S. National Institutes of Health (NIH). Los Alamos National Laboratory (LANL), which is operated by the University of California for the Department of Energy, is located in Los Alamos, New Mexico. LANL gathers, annotates, and organizes the database and transmits it to BBN Laboratories Incorporated, a research and consulting firm in Cambridge, Massachusetts. The collected information is prepared for release by BBN and distributed to subscribing institutions and scientists in regular updates. Cosponsors of the GenBank project include the National Cancer Institute, the National Institute of Allergy and Infectious Diseases, the National Library of Medicine, the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, and the Division of Research Resources (DRR) of NIH, as well as the National Science Foundation, the U.S. Department of Energy, and the U.S. Department of Defense. GenBank's first release was in October 1982.

## 1.3 New features of this edition

- . The Citation Index has been added to assist readers in finding bibliographical citations for journal articles. This new index lists journal title, volume number, page numbers, and year of publication for each article cited.
- . As a result of limited resources and an ever-increasing rate of sequence publication, it has not been possible to collect and present all sequences in the fully annotated form that we would like. It is nevertheless vitally important

that at least as much raw sequence data as possible be presented. Therefore, we have a new section entitled Unannotated Sequences, which contains unannotated and unclassified sequences and citations. We hope that in the future we will have the resources to move this information rapidly into its proper position in the main database.

A separate volume is now available that contains master indices for the entire database as well as a master directory for all of the entries in the database.

## 2. Contents of the compendium

As combined in this compendium, the two databases contain a total of nearly 8.5 million bases from 6700 articles. The following indices are provided to assist users in finding the information they need: the Keyword Phrase Index, the Taxonomic Classification Index, the Author Index, the Citation Index, the Accession Number Index, the EMBL Entry Index, and the GenBank Entry Index. Most of the entries are annotated to indicate the locations within the reported sequences of coding regions and other experimentally determined sites of biological significance. Full bibliographic information is included in every entry, and many of the entries also include comments abstracted from the original papers. Technical appendices located after the main data sections in each volume contain detailed explanations of information in the entries.

### 2.1 General organization of the compendium

The entries in the compendium are presented in thirteen sections; within each section the entries are grouped according to the source organism. These sections are arranged in eight volumes, as follows:

#### Volume I. Primates

##### Section 1. Primate Sequences

#### Volume II. Rodents

##### Section 2. Rodent Sequences

#### Volume III. Other Vertebrates and Invertebrates

- Section 3. Other Mammalian Sequences
- Section 4. Other Vertebrate Sequences
- Section 5. Invertebrate Sequences

#### Volume IV. Plants and Organelles

- Section 6. Plant Sequences
- Section 7. Organelle Sequences

#### Volume V. Bacteria and Bacteriophage

- Section 8. Bacterial Sequences
- Section 9. Bacteriophage Sequences

#### Volume VI. Viruses

##### Section 10. Viral Sequences

#### Volume VII. Structural RNA, Synthetic, and Unannotated Sequences

- Section 11. Structural RNA Sequences
- Section 12. Synthetic Sequences
- Section 13. Unannotated Sequences

#### Volume VIII. Database Directory and Master Indices

Each volume of the compendium contains this introduction, one or more sections of data, technical appendices, and indices to that volume. The Author Index, the Citation Index, the Taxonomic Classification

Index, the Keyword Phrase Index, the Accession Number Index, the EMBL Entry Index, and the GenBank Entry Index in Volume VIII are master indices to all of the volumes in this edition.

### 2.2 Finding an entry

Users approaching the database for the first time must determine which section contains the sequence they are looking for. Most of the sections are self-explanatory, but it is helpful to point out the following conventions:

- Yeast and fungal sequences are in the Plant Sequences section.
- Plasmids and transposons isolated from bacteria are listed in the Bacterial Sequences section.
- The Structural RNA section includes the sequences of mature transfer RNA, ribosomal RNA, small nuclear RNA, and other structural RNA molecules. All structural RNA genes and most structural RNA precursor sequences are listed with their organisms in their particular sections.
- The Synthetic Sequences section includes any nucleic acid sequence that is created in a laboratory and does not occur naturally, including synthetic plasmids that are not included with the other bacterial sequences. The major exceptions to this rule are cDNA sequences, since they are regarded as a means of sequencing naturally occurring RNA sequences.

The individual entries within each section are arranged alphabetically by entry name. Summary tables and section directories are included at the beginning of each section to provide some guidance for locating the entries. Table 1 is an overall summary table of the entire database. This table shows the names of the sections, as well as the numbers of reported sequences, distinct entries, and nucleotide bases in each section. There are typically more reported sequences than entries because overlapping sequences are frequently merged into a single, combined entry.

A table that summarizes the entries appears at the beginning of each section. This table is called the Section Summary. The Section Summary for the Primate Sequences section, for example, lists, by organism (e.g., Ape), the corresponding organism code (e.g., APE), the number of reported sequences for that organism, the number of entries, the number of bases, and the page number on which this group of entries begins.

Note that the page numbers throughout are arranged separately for each section. The numbers are printed on each page with a short section prefix. For example, the first three pages of Section 1: Primate Sequences are numbered PRIMATE-1, PRIMATE-2, and PRIMATE-3. Table 1 shows the page number prefix for each section.

A detailed alphabetized directory for the section appears immediately after the Section Summary. The section directory contains one line of information for each entry in the section and serves as a complete table of contents for that section, listing the full entry name, the description and length of each entry (i.e., number of base pairs), and the page on which each entry appears.

### 3. How to Read an Entry

The entries for each section begin after the section directory. Each entry is separated from the next by a dashed line running the width of the page. There are two types of entries in the compendium: (1) self-contained, and (2) segmented. Segmented entries are used when noncontiguous pieces of the same nucleic acid molecule have been sequenced and the ordering of the pieces is known.

Table 1: Summary of Sequences Presented in Each Section

Section Number	Section Code	Section Description	Number of Sequences	Number of Entries	Number of Bases
1	PRIMATE	Primate Sequences	1492	1028	1240779
2	RODENT	Rodent Sequences	1638	1272	1111622
3	MAMMAL	Other Mammalian Sequences	293	245	244554
4	VERT	Other Vertebrate Sequences	557	474	400509
5	INVERT	Invertebrate Sequences	696	605	435280
6	PLANT	Plant Sequences	717	594	643365
7	ORGANELLE	Organelle Sequences	434	368	485666
8	BACT	Bacterial Sequences	1130	749	1031546
9	PHAGE	Bacteriophage Sequences	338	160	271817
10	VIRAL	Viral Sequences	1748	1093	1517025
11	RNA	Structural RNA Sequences	734	637	69232
12	SYNTHETIC	Synthetic Sequences	259	224	72029
13	UNANNOTATED	Unannotated Sequences	1377	1374	918933
Overall Summary:			11413	8823	8442357

### 3.1 Summary of the entry fields

Each entry is composed of several kinds of information, referred to here as fields. Not every field appears in every entry, but the full list of possible fields, in the order in which they appear, is as follows:

- Entry Name - a short, unique name providing the label for the entry.
- Definition - a brief description of the sequence, beginning with the name of the source organism.
- Segment - indicates which segment this entry is in a series of separated sequences from the same molecule.
- EMBL ID - entry name(s) in the EMBL database that correspond to the entry names in this work.
- Accession Numbers - short codes that provide unique, unchanging identifiers for the data in each entry; the first number in the list is known as the primary accession number of the entry.
- Date - the year, month, and day when this form of the entry appeared in the GenBank version of the database, plus information on whether the entry is preliminary or complete.
- References - citations for all references used to construct each entry.
- Keywords - short phrases describing gene products and other information pertinent to looking up an entry.
- Source - most commonly used name of the source organism, followed by a formal scientific name.
- Comment - information that does not readily fall into the other fields, including information abstracted from the references and cross-references to other entries.
- Features and Sites Tables - tables designed to describe locations and regions of biological significance within the sequence.
- Origin - describes the start of a sequence in relation to an experimentally determined site.
- Sequence - statistics on the numbers and kinds of bases in the sequence, followed by the sequence itself.

See Example 1 for an example of a typical pair of entries.

### 3.2 The fields in detail

#### ENTRY NAME

##### EMBL "ID" Names and GenBank "Locus" Names

The entry name is a short, unique name that provides the label for an entry. In order to organize this compendium in a coherent fashion, it was necessary to choose a uniform method for naming all of the entries, regardless of which database the information was extracted from. By mutual agreement, we have presented the entries under the names assigned to them in the GenBank database. The conventions for choosing these names, which include abbreviations for the organisms from which the nucleic acids were isolated, are described in detail in Technical Appendix A: Entry Name and Molecule Type Conventions.

The GenBank entry names have been called "locus names" throughout this book, and there are many occasions where one entry refers to another "locus" or another group of "loci"; this terminology is simply a way of referring to other entries. The entry names used for the corresponding information in the EMBL Sequence Library are given after the label "EMBL ID:" in the second line of each entry. Not all entries have been assigned EMBL ID names at this stage of our collaboration, but eventually all entries will be assigned names in both databases, and we are actively moving toward a common naming system for corresponding entries in the two databases.

The GenBank Entry Index lists all of the GenBank entry names alphabetically, together with the section name and page number on which the entry begins. The other indices refer to GenBank entry names, not page numbers, since these are the names used in organizing the book. The page numbers must be looked up in the GenBank Entry Index.

#### DEFINITION

The definition of an entry provides a brief description of the sequence. This definition is used to construct the listing for the entry in the section directory. Typically it includes the name of the organism and other important information describing the entry. Information about the type of molecule and whether the sequence presented is circular or a complete tandem repeat is included in brackets at the end of the definition for most entries. The conventions used in specifying the molecule type are described in detail in Technical Appendix A: Entry Name and Molecule Type Conventions.

## INTRODUCTION

```

ANIMTCYB1: a.nidulans mt apocytochrome b (coba) gene; exon1. [DNA] SEGMENT: 1 of 2
EMBL ID: M1AN02 ACCESSION NUMBERS: J01388 V00651 DATE: updated 83-11-01
REFERENCES: [1] (bases 1 to 838) Waring,R.B., Davies,R.W., Lee,S., Grisi,E., Berks,M.M. and Scazzocchio,C.; "the
mosaic organization of the apocytochrome b gene of aspergillus nidulans revealed by dna sequencing"; Cell
27, 4-11 (1981)
KEYWORDS: cytochrome; apocytochrome.
SOURCE: aspergillus nidulans. Mitochondrion Aspergillus nidulans
COMMENT: Single intron of about 1050 bp occupies same position as I3 in "long" S. cerevisiae gene. Open reading
frame of exon 1 continues at least 200bp into ivs. TGA codes for trp. See <hummt> and <ystmtcyb>. See
other loci beginning <animtcyb>.

SITES:
key site span description FEATURES:
key from to description
refnumbr 1 1 numbered -125 in [1]; zero not used. key 126 + 631 apocytochrome b (exon 1)
->pept 126 1 coba coding sequence start FEATURES:
pept/IVS 632 0 coba ivs1 start (exon1 end) key from to description
CDS 126 631 apocytochrome b part 1 (631 is
2nd base in codon)
IVS 632 >838 intron I

ORIGIN: near hind iii site in bgl ii fragment 4.
SEQUENCE: 838 bp 320 a 112 c 132 g 274 t
1 atataaaaca gtaattaata aataaaataa ttactttaat cttagagatt ttaaactcga taaataaaaa aaaaaaata aataaataaa ttaagtgaag
101 aaaaaaaaaa aaaaaaaaaa aaaaaatgag aattttaaaa agtcatcctt tactaaaaat agtaaattcg tatataatag attcacctca accagctaag
201 ttaagttatt tatgaaattt cggatcatta ttagctttat gtttaggtat acaaaatagta acagggtgta cattagctat gcattatata cctagtgtat
301 cagaagcatt taattctgta gagcatatta tgagagatgt aaataatcca tgattagtag gttacttaca ctctaataca gottcagctt tcttctttt
401 agtatactta cacataggaa gaggtttata ttatggatct tacaaaaacac ctagaacttt aacatgagct attggaacag taataactaat agttatgatg
501 gccacagcct tcttaggtta tgttttacct tatggtcaaa tgagttttat aggtgtaca gttattacta acctaatgag tgctatacct tgaataggtc
601 aagatatatt tgagtttatt tgaggaggtt tatacacaga tgaaccacaa tgcggtgacg tattgtttaa aatcctgctt aatgctggaa aatccccaat
701 cttaggattt gcatacgact tattctttat aatagtatta ttaataggog tgaaaatgac aatgacacgg ggaaaatcag cagggttgag aagtttacat
801 acttcagaag cctctcagag actacatgca ggagatct

ANIMTCYB2: a.nidulans mt apocytochrome b (coba) gene; exon2. [DNA] SEGMENT: 2 of 2
EMBL ID: M1AN03 ACCESSION NUMBERS: J01389 V00652 DATE: updated 83-11-01
REFERENCES: [1] (bases 1 to 1082) Waring,R.B., Davies,R.W., Lee,S., Grisi,E., Berks,M.M. and Scazzocchio,C.; "the
mosaic organization of the apocytochrome b gene of aspergillus nidulans revealed by dna sequencing"; Cell
27, 4-11 (1981)
KEYWORDS: cytochrome; apocytochrome.
SOURCE: aspergillus nidulans. Mitochondrion Aspergillus nidulans
COMMENT: Single intron of about 1050 bp occupies same position as I3 in "long" S. cerevisiae gene. Open reading
frame of exon 1 continues at least 200bp into ivs. TGA codes for trp. See <hummt> and <ystmtcyb>. See
other loci beginning <animtcyb>.

SITES:
key site span description FEATURES:
key from to description
IVS/pept 77 0 coba exon2 start (ivs1 end) key + 77 734 apocytochrome b (exon 2)
pept<- 734 1 coba coding sequence end FEATURES:
key from to description
CDS 77 731 apocytochrome b part 2 (77 is
3rd base in codon)
IVS <1 76 intron I

ORIGIN: about 750 bp after animtcyb1
SEQUENCE: 1082 bp 373 a 123 c 140 g 446 t
1 gatcaataaa gaaatttatt gcgtatagta agaggattta atatttata taaatctgta actatcaaca taaatgctct gtaaataatg caactttaaa
101 cagattcttt gcattacatt tcttattacc ttittgtatta gctgctttag cattaatgca ttaataagct atgcatgata cagtaggatac aggtaatcct
201 ttaggtattt ctgctaatta cgatagatta ccttttgctc ctatttttat atttaaagat ttaataacta tattttattt cttttattga ttatcaatat
301 ttgttttctt tatgcctaatt gctttagggtg atagtgaata ttatgttatg gctaactcta tgcaaacctc acctgctata gttccagaat gatattcttt
401 acctttctat gctatttttaa gatctatacc taataaatta ttaggtgtta tagctatggt tgcgtctata ttagcattaa tgggtatgcc tataactgat
501 ttatctaaat taagaggagt acaatttaga cctttaagta aagtagtatt ctatattttt gtactgaact tcttaatat aatgcaaaata ggtgcaaaac
601 acgttgaac tccatttatt gaatttggac aaatttctac tattatttat ttgcatatt tctttgtaat agttcctgtt gttagtttaa ttgaaaatac
701 tttagtagaa ttaggaacta aaaaaaactt ttaattctta gtccctctag gaaaaaaa caaatttatt aaacagctgc aaatttaatt tatgaaaatg
801 atattagaca aaaaatttta aaagaatta gatagctaca ttgtattata atcaatttat taatatattt gttttcatct atactttgta gttaatcata
901 agtatgatgt aataaatagt aatatctttt taaagtagac ttgaccttta aaatttttaa tataattatt attatcttgt tagagtataa ttataataca
1001 tataatattg tatattagga gtttagagca atgggtttgc gttttgattg caaattgaaa tatagggtat cgatttcccc gg

```

Example 1. Two segmented entries from the Organelle Sequences section.



## SEGMENT

In those cases where an entry is segmented, a segment field is used to indicate which segment this entry is in a series of noncontiguous sequences from the same molecule. Segmented entries contain a label after the molecule type, at the end of the first line of the entry, which indicates the position of this entry in the group of segmented entries. The number at the end of the entry name also indicates which segment of a complete sequence is contained in this entry.

## ACCESSION NUMBERS

Accession numbers have been assigned to all entries in the EMBL Sequence Library and the GenBank database, using a system that was worked out jointly by the two database teams. These arbitrary labels for the data included in an entry consist of a single letter followed by five digits; unlike the entry names, they carry no information about the type or nature of the entry. Accession numbers never change, but rather follow along with the data they point to no matter how the entry or entries in question might be reorganized in future versions of either database. For example, if two different entries with different accession numbers are merged into a single entry, both accession numbers are included in the new entry.

The first number in the list of accession numbers is the primary accession number for the entry. If you cite information contained in an entry in either database, you are encouraged to include the primary accession number in your citation, since this number will enable your readers to find the data in question in future releases of either database. The Accession Number Index lists all accession numbers that have been assigned to date and the entries to which they have been assigned.

## DATE

The date field contains a date in the form year-month-day, preceded by the word "entered", "updated", "pre-entry", or "unannotated". The date given is the date of the most recent GenBank release in which this entry underwent any major revisions. If the word "entered" appears before the date, this means that this entry was first entered in the database on the date given, and that it has not undergone any substantial revisions since that release. If there has been some substantial revision, such as the addition of another reference, the date of the most recent revision is given preceded by the word "updated." If the word "pre-entry" appears, it indicates that although the complete sequence and some portion of the annotations appear in this entry, there are additional annotations that will be included once review of the article in question is completed. Pre-entries sometimes undergo several rounds of updates and revisions before they are upgraded to full entries. The date field in all of the entries in the Unannotated Sequences section begins with the word "unannotated".

## REFERENCES

This field includes the number (in brackets) assigned to each cited paper; a brief statement of which information in the entry comes from each reference (shown in parentheses); and the actual citation of the article. Sequence data submitted directly to the EMBL or GenBank databases and not published elsewhere are listed as unpublished references. These references generally have no title; they simply list the contributor as the author and include the word "Unpublished" followed by the year in parentheses and the address of the contributor. See Technical Appendix B: Reference Citation Conventions for further information on reference format.

## KEYWORDS

The keywords field contains short words or phrases that identify gene products and other useful identifying characteristics of the entries. The Keyword Phrase Index provides a means for looking up all entries that have common keyword phrases; if a particular entry has no keyword phrases, this indicates that none of the phrases in the keyword index apply to it. A small number of pre-entries show the word "unassigned" in place of any keyword phrases; this word indicates that the entry has not yet been reviewed to determine which, if any, of the keyword phrases apply to it.

## SOURCE

The source field is an English language statement about the organism and tissue from which the sequence was isolated. If the entry contains a viral sequence, the host organism is usually also named. This statement is followed by a formal designation of the source organism, typically consisting of the full scientific name. The Taxonomic Classification Index lists all entries according to their formal taxonomic classifications.

## COMMENT

The comment includes information that does not readily fall in other fields, such as statements abstracted from the references and cross-references to other entries.

## FEATURES AND SITES TABLES

Three different tables can appear in each compendium entry: the EMBL features table, the GenBank features table, and the sites table. All three are designed to describe regions and locations of biological significance within the sequence. The two features tables show regions, with starting and ending points for each feature of interest. The sites table, on the other hand, shows individual locations of interest within the sequence, together with a number that indicates whether the location is a single point or encompasses multiple bases. The features tables presented come from both databases; those with lowercase keywords in the key column are in GenBank format, while those with uppercase keywords are in EMBL format. The EMBL features tables are included in those cases where the numbering system used in the EMBL entry corresponds to the numbering used in the GenBank entry and the EMBL table provides some information that augments the information found in the other two tables. There is often considerable redundancy in the information contained in these three tables. Full information about the conventions used in constructing these tables appears in Technical Appendix C: Sites and Features Tables.

## ORIGIN

The origin field describes the start of the sequence (the 5' terminus) in relation to some experimentally determined site, such as a restriction enzyme cutting site.

## SEQUENCE

The first line in the sequence field gives the total number of base pairs, and the number of adenines, cytosines, guanines, thymines (or uracils), and other bases reported in the sequence. After this information, the actual sequence is listed. The bases in the sequence are numbered line by line, and they are presented one hundred bases per line, in groups of ten.

## 4. Two Sample Entries

Example 1 is a reproduction of a typical pair of segmented entries found in the Organelle Sequences section. In this example, the first entry consists of the 5' portion of a particular sequence and the second consists of the 3' portion of the same sequence. The two entries in this excerpt have only been partly converted to full mixed-case representation. When there are free text portions of an entry that are not converted to mixed case, they are shown in lower-case characters. The common prefix for the two entry names is ANIMTCYB; therefore, the first entry is called ANIMTCYB1 and the second ANIMTCYB2. The segment field at the end of the first line of each of these entries states explicitly that the first entry is the first of two segments and the next is the second of two segments.

The entry names in this example consist of several parts. The first three characters in both names are "ANI", an abbreviation for the fungus *Aspergillus nidulans*. The letters "MT" indicate that these are mitochondrial sequence entries, and "CYB" is an abbreviation for apocytochrome b. The last character in each entry name is the segment number. See Technical Appendix A for a description of the conventions and abbreviations used in assigning entry names.

The definition following the entry name gives the name of the source organism and other information describing the sequence. For example, "a. nidulans mt apocytochrome b (coba) gene; exon 1" indicates that the entry called ANIMTCYB1 contains exon 1 of the mitochondrialoba gene coding for apocytochrome b in *Aspergillus nidulans*. The notation "[DNA]" immediately following the definition indicates that the sequence represents a double-stranded DNA molecule.

On the second line of each entry, the corresponding EMBL ID names are given. The entry called ANIMTCYB1 corresponds to the EMBL entry called M1AN02, and ANIMTCYB2 corresponds to M1AN03. The accession numbers appear next. Each of these entries has two accession numbers, since these particular entries were originally entered independently in each database. The date field at the end of the second line indicates that these entries were most recently revised in the GenBank release dated 1 November 1983.

The list of references begins on the third line of these entries. The notation "[1] (bases 1 to 838)" indicates that the reference in the first entry is referred to as reference [1] and that it is the source of the bases numbered from 1 to 838 in the entry. The remainder of the reference listing is a fairly conventional citation for the particular reference.

The keywords field appears next. These two entries can be looked up using the two keywords "cytochrome" and "apocytochrome" in the Keyword Phrase Index. The next field, the source, lists first the commonly used name of the source organism, *Aspergillus nidulans*, followed by the scientific name used to classify the organism in the taxonomic classification index, Mitochondrion *Aspergillus nidulans*. In many cases, such as those in this example, the two parts of the source field are somewhat redundant.

The comment, beginning on the next line of the entry, gives brief information about the entry (taken from the reference) and refers the reader to other loci (entries) that have names beginning "ANIMTCYB".

Three tables appear next in each of these entries: a sites table and two features tables. All three of these tables indicate the portions of these sequences that code for apocytochrome b; this includes the bases numbered from 126 to 131 in ANIMTCYB1 and bases 77 to 734 in ANIMTCYB2. The EMBL features tables were included in these two entries since they explicitly state the reading frame for the codons in these

regions. These two entries also illustrate the use of some of the special symbols that can appear in both styles of features tables. The "+" signs in the GenBank features tables indicate that the coding region resumes in another segment. The "<" and ">" characters in the EMBL features tables indicate that the intervening sequence extends beyond the ends of the reported sequences. The apparent differences in the numbers reported in the three tables in these entries are due to systematic differences in the conventions that have been used in the EMBL and GenBank databases. For example, the GenBank features table reports that the coding sequence in ANIMTCYB2 terminates at base 734, while the corresponding EMBL features table reports the end of the coding region as base 731. This apparent difference merely reflects the GenBank convention of including the termination codon in the reported coding region and the EMBL convention of excluding the termination codon.

The origin lines, which follow the features and sites tables in these entries, indicate that the sequence in ANIMTCYB1 starts near a particular restriction enzyme cut site and that the sequence in ANIMTCYB2 is separated from that in ANIMTCYB1 by approximately 750 bases.

The sequence field is the last field in the entry. The first line of the first entry shows that there are 838 base pairs in this sequence, which includes 320 adenines, 112 cytosines, 132 guanines, and 274 thymines. Finally, the actual sequence is listed and clearly numbered.

Please refer to the technical appendices after the data sections in each volume for additional details about the conventions used in presenting the entries in this collection. The indices at the end of each volume include brief explanations and instructions for their use.

## Section 1: Primate Sequences

Entry Name	Description and Length	Page
APEERVMB	Ape (baboon) endogenous retrovirus DNA homologous to the polymerase gene of M-MuLV. 173BP	PRIMATE-14
APEERVMC	Ape (chimpanzee) endogenous retrovirus DNA homologous to the polymerase gene of 192BP	PRIMATE-14
APEHBA1M	Ape (chimpanzee) alpha-1-globin mRNA. 551BP	PRIMATE-14
APEHBA2M	Ape (chimpanzee) alpha-2-globin mRNA. 549BP	PRIMATE-15
APEHBAPS	Chimpanzee (P.troglodytes) pseudo alpha-globin gene, 5'-flank (Alu repeat). 1076BP	PRIMATE-15
APEHBB3CH	Ape (Chimpanzee) beta-globin mRNA, 3' untransl. region. 136BP	PRIMATE-15
APEHBB5CH	ape (chimpanzee) beta-globin mrna, 5' untransl. region. 54BP	PRIMATE-16
APEHBBPCH	Chimpanzee beta-globin pseudogene. 2147BP	PRIMATE-16
APEHBBPG	Gorilla beta-globin pseudogene. 2151BP	PRIMATE-16
APEIL2	Ape (gibbon) interleukin 2 mRNA. 730BP	PRIMATE-16
APEIL2LTR	Ape (gibbon) interleukin 2 mRNA containing inserted LTR sequence of gibbon ape 1061BP	PRIMATE-17
APERSA	Chimpanzee Alu type DNA. 3163BP	PRIMATE-17
GCRSA2A	Galago crassicaudatus (bush baby) Alu family type II, clone GAL7. 245BP	PRIMATE-18
GCRSA2B	G.crassicaudatus (bush baby) Alu family type II, clone GAL 16. 179BP	PRIMATE-18
GCRSA2C	G.crassicaudatus (bush baby) Alu family type II, clone GAL 5. 292BP	PRIMATE-18
GCRSA2D	G.crassicaudatus (bush baby) Alu family type II, clone GAL 6. 179BP	PRIMATE-18
GCRSA2E	G.crassicaudatus (bush baby) Alu family type II, clone GAL 12. 241BP	PRIMATE-18
GCRSA2F	G.crassicaudatus (bush baby) Alu family type II, clone GAL 25. 226BP	PRIMATE-19
GCRSA2G	G.crassicaudatus (bush baby) Alu family type II, clone GAL 33. 233BP	PRIMATE-19
GCRSA2H	G.crassicaudatus (bush baby) Alu family type II, clone GAL 34. 230BP	PRIMATE-19
GCRSA2I	G.crassicaudatus (bush baby) Alu family type II, clone GAL 27. 264BP	PRIMATE-19
GCRSA2J	G.crassicaudatus (bush baby) Alu family type II, clone GAL 3. 146BP	PRIMATE-19
GCRSA2K	G.crassicaudatus (bush baby) Alu family type II, clone GAL 30. 238BP	PRIMATE-20
GCRSA2L	G.crassicaudatus (bush baby) Alu family type II, clone GAL 21. 192BP	PRIMATE-20
GCRSA2M	G.crassicaudatus (bush baby) Alu family type II, clone GAL 1. 183BP	PRIMATE-20
GCRSA2N	G.crassicaudatus (bush baby) Alu family type II, clone GAL 20. 280BP	PRIMATE-20
GCRSA2O	G.crassicaudatus (bush baby) Alu family type II, clone GAL 26. 246BP	PRIMATE-20
GCRSA2P	G.crassicaudatus (bush baby) Alu family type II, clone GAL1 4. 199BP	PRIMATE-20
GCRSA2Q	G.crassicaudatus (bush baby) Alu family type II, clone GAL 35. 245BP	PRIMATE-21
GCRSA2R	G.crassicaudatus (bush baby) Alu family type II, clone GAL 40. 258BP	PRIMATE-21
GCRSA2S	G.crassicaudatus (bush baby) Alu family type II, clone GAL 39. 107BP	PRIMATE-21
HUM7SLR1	Human 7SL RNA pseudogene, clone p7L30.1. 377BP	PRIMATE-21
HUM7SLR2	Human 7SL RNA pseudogene, clone p7L30.2. 378BP	PRIMATE-21
HUM7SLRA	Human 7SL RNA pseudogene, clone p7L28. 313BP	PRIMATE-22
HUM7SLRB	Human 7SL RNA pseudogene, clone p7LEMI. 313BP	PRIMATE-22
HUM7SLRC	Human 7SL RNA pseudogene, clone p7L7. 345BP	PRIMATE-22
HUM7SLRD	Human 7SL RNA pseudogene, clone p7L23. 384BP	PRIMATE-22
HUM7SLRE	Human 7SL RNA pseudogene, clone p7L63. 285BP	PRIMATE-23
HUMA1ACM	Human alpha-1-antichymotrypsin complete gene, mRNA. 1520BP	PRIMATE-23
HUMA1ACMA	Human alpha-1 antichymotrypsin mRNA, 3' end. 802BP	PRIMATE-23
HUMA1AT1	Human alpha-1-antitrypsin gene: 5' terminus. 384BP	PRIMATE-24
HUMA1AT2	Human alpha-1-antitrypsin gene: exon 1 (partial). 101BP	PRIMATE-24
HUMA1AT3	Human alpha-1-antitrypsin gene: ivs a/ exon 2 junction. 69BP	PRIMATE-24
HUMA1AT4	Human alpha-1-antitrypsin gene: 3' terminus. 292BP	PRIMATE-25
HUMA1ATM	Human alpha-1-antitrypsin mRNA, complete cds. 1352BP	PRIMATE-25
HUMA1ATP	Human alpha-1-antitrypsin gene (S variant), complete cds. 12222BP	PRIMATE-26
HUMA1ATR	Human alpha-1-antitrypsin mRNA, complete cds. 1346BP	PRIMATE-28
HUMA1ATS	Human alpha-1-antitrypsin mRNA, carboxyterminal region. 460BP	PRIMATE-28
HUMA2TPI	Human alpha-2-thiol proteinase inhibitor mRNA, complete coding sequence. 1493BP	PRIMATE-28
HUMACBPA1	Human cytoplasmic beta-actin related pseudogene H-beta-Ac-psi-1, complete cds. 1660BP	PRIMATE-29
HUMACBPA2	Human cytoplasmic beta-actin related pseudogene H-beta-Ac-psi-1, 3'end. 192BP	PRIMATE-29
HUMACBPB1	Human cytoplasmic beta-actin related pseudogene H-beta-Ac-psi-2, complete cds. 1665BP	PRIMATE-30
HUMACBPB2	Human cytoplasmic beta-actin related pseudogene H-beta-Ac-psi-2, 3' end. 196BP	PRIMATE-30
HUMACCYBA	Human cytoplasmic beta-actin gene, complete cds. 3657BP	PRIMATE-30
HUMACCYBB	Human cytoplasmic beta-actin gene, complete cds. 3646BP	PRIMATE-31
HUMACHRA1	Human acetylcholine receptor alpha-subunit gene, exon P1. 198BP	PRIMATE-32
HUMACHRA2	Human acetylcholine receptor alpha-subunit gene, exons P2 and P3. 513BP	PRIMATE-32
HUMACHRA3	Human acetylcholine receptor alpha-subunit gene, exon P4. 229BP	PRIMATE-33
HUMACHRA4	Human acetylcholine receptor alpha-subunit gene, exon 5. 363BP	PRIMATE-33
HUMACHRA5	Human acetylcholine receptor alpha-subunit gene, exon P6. 417BP	PRIMATE-33
HUMACHRA6	Human acetylcholine receptor alpha-subunit gene, exon P7. 309BP	PRIMATE-33
HUMACHRA7	Human acetylcholine receptor alpha-subunit gene, exons P8 and P9. 2318BP	PRIMATE-34
HUMACHRG1	Human acetylcholine receptor gamma subunit gene, exons 1 and 2. 709BP	PRIMATE-34
HUMACHRG2	Human acetylcholine receptor gamma subunit gene, exons 3 and 4. 480BP	PRIMATE-35
HUMACHRG3	Human acetylcholine receptor gamma subunit gene, exon 5. 240BP	PRIMATE-35
HUMACHRG4	Human acetylcholine receptor gamma subunit gene, exon 6. 435BP	PRIMATE-35
HUMACHRG5	Human acetylcholine receptor gamma subunit gene, exons 7 and 8. 714BP	PRIMATE-36
HUMACHRG6	Human acetylcholine receptor gamma subunit gene, exon 9. 344BP	PRIMATE-36
HUMACHRG7	Human acetylcholine receptor gamma subunit gene, exons 10 and 11. 690BP	PRIMATE-36
HUMACHRG8	Human acetylcholine receptor gamma subunit gene, exon 12. 703BP	PRIMATE-37
HUMACTASK	Human adult skeletal muscle alpha-actin mrna. 1374BP	PRIMATE-37

# **DATABASE DIRECTORY**

Entry Name	Description and Length	Page
HUMACTBET	Human fibroblast beta-actin mrna, 5' end. 57BP	PRIMATE-38
HUMACTCA1	Human alpha-cardiac actin gene, 5' flank and exon 1. 232BP	PRIMATE-38
HUMACTCA2	Human alpha-cardiac actin gene, exons 2, 3, and 4. 1846BP	PRIMATE-38
HUMACTCA3	Human alpha-cardiac actin gene, exon 5. 442BP	PRIMATE-39
HUMACTCA4	Human alpha-cardiac actin gene, exon 6 and 3' flank. 749BP	PRIMATE-39
HUMACTFIB	Human actin gene from fibroblast. 83BP	PRIMATE-40
HUMACTGAM	Human fibroblast gamma-actin mrna, 5' end. 57BP	PRIMATE-40
HUMACTSM1	Human smooth muscle actin gene (aortic): exon 1 & ivs flanks. 976BP	PRIMATE-40
HUMACTSM2	Human smooth muscle actin gene (aortic): exon 2 and ivs flanks. 292BP	PRIMATE-41
HUMACTSM3	Human smooth muscle actin gene (aortic): exon 3 and ivs flanks. 223BP	PRIMATE-41
HUMACTSM4	Human smooth muscle actin gene (aortic): exon 4 and ivs flanks. 203BP	PRIMATE-41
HUMACTSM5	Human smooth muscle actin gene (aortic): exon 5 and ivs flanks. 530BP	PRIMATE-41
HUMACTSM6	Human smooth muscle actin gene (aortic): exon 6 and ivs flanks. 269BP	PRIMATE-42
HUMACTSM7	Human smooth muscle actin gene (aortic): exon 7 and ivs flanks. 408BP	PRIMATE-42
HUMADA	Human adenosine deaminase mRNA, complete cds. 1478BP	PRIMATE-42
HUMADAM	Human adenosine deaminase (ADA-1) mRNA, partial: 3' 2/3 of coding. 982BP	PRIMATE-43
HUMADAM2	Human adenosine deaminase mRNA, complete coding sequence. 1535BP	PRIMATE-43
HUMADHIB1	Human class I alcohol dehydrogenase beta subunit gene, exon 1. 375BP	PRIMATE-43
HUMADHIB2	Human class I alcohol dehydrogenase beta subunit gene, exon 2. 132BP	PRIMATE-44
HUMADHIB3	Human class I alcohol dehydrogenase beta subunit gene, exon 3. 169BP	PRIMATE-44
HUMADHIB4	Human class I alcohol dehydrogenase beta subunit gene, exon 4. 118BP	PRIMATE-44
HUMADHIB5	Human class I alcohol dehydrogenase beta subunit gene, exon 5. 250BP	PRIMATE-45
HUMADHIB6	Human class I alcohol dehydrogenase beta subunit gene, exon 6. 291BP	PRIMATE-45
HUMADHIB7	Human class I alcohol dehydrogenase beta subunit gene, exon 7. 166BP	PRIMATE-45
HUMADHIB8	Human class I alcohol dehydrogenase beta subunit gene, exon 8. 169BP	PRIMATE-46
HUMADHIB9	Human class I alcohol dehydrogenase beta subunit gene, exon 9. 700BP	PRIMATE-46
HUMAFH	Human apoferritin (H chain) mRNA. 801BP	PRIMATE-47
HUMALB	Human albumin mRNA, complete cds. 614BP	PRIMATE-47
HUMALBA	Human albumin mRNA, complete cds. 615BP	PRIMATE-47
HUMALBAF1	Human serum albumin mRNA, complete cds. 1929BP	PRIMATE-48
HUMALBAF2	Human serum albumin gene, 3' flank. 336BP	PRIMATE-49
HUMALBAF3	Human DNA between albumin and alpha-fetoprotein gene. 87BP	PRIMATE-49
HUMALBAF4	Human alpha-fetoprotein (AFP) mRNA, complete cds. 2032BP	PRIMATE-49
HUMALBG	Human serum prealbumin gene. 7619BP	PRIMATE-50
HUMALDB	Human aldolase B complete coding region, mRNA. 1491BP	PRIMATE-51
HUMALDBX	Human aldolase B mRNA, complete cds. 1595BP	PRIMATE-52
HUMALDH1	Human aldehyde dehydrogenase 1 mRNA. 1560BP	PRIMATE-52
HUMALDH2	Human aldehyde dehydrogenase 2 mRNA. 1603BP	PRIMATE-53
HUMALPA	Human alpha repeat sequence dna. 340BP	PRIMATE-53
HUMAMYAP	Human pancreatic alpha-amylase mRNA. 1566BP	PRIMATE-53
HUMAMYAS	Human salivary gland alpha-amylase mRNA. 1770BP	PRIMATE-54
HUMANFA	Human PND gene encoding atrial natriuretic factor, complete cds. 2710BP	PRIMATE-54
HUMANFB	Human prepronatriuretic factor, atrial natriuretic factor coding region. 108BP	PRIMATE-55
HUMANFH	Human atrial natriuretic factor, partial cds. 345BP	PRIMATE-56
HUMANG	Human angiotensinogen mRNA, complete CDS. 2099BP	PRIMATE-56
HUMAPOAI1	Human apolipoprotein A-I and C-III genes, complete cds. 8337BP	PRIMATE-57
HUMAPOAI2	Human apolipoprotein A-IV mRNA (partial). 954BP	PRIMATE-59
HUMAPOAI3	Human apolipoprotein gene A-II on chromosome 1. 2956BP	PRIMATE-60
HUMAPOB	Human apolipoprotein B100 mRNA (partial cds). 4634BP	PRIMATE-61
HUMAPOBX	Human apolipoprotein B mRNA, partial cds. 593BP	PRIMATE-62
HUMAPOCI	Human apolipoprotein C-I mRNA. 419BP	PRIMATE-62
HUMAPOCII	Human apolipoprotein C-II gene, complete cds. 4340BP	PRIMATE-63
HUMAPOE3	Human apolipoprotein E (epsilon 2 and 3 alleles) mRNA. 1156BP	PRIMATE-64
HUMAPOE4	Human apolipoprotein E (epsilon-4 allele) gene, complete cds. 5515BP	PRIMATE-65
HUMARS1A	Human autonomously replicating sequence 1 (ARS1). 456BP	PRIMATE-66
HUMARS2A	Human autonomously replicating sequence 2 (ARS2). 1600BP	PRIMATE-67
HUMAS1	Human argininosuccinate synthetase gene, exon 1 and flanks. 198BP	PRIMATE-67
HUMAS1PS	Human argininosuccinate synthetase pseudogene 1. 1877BP	PRIMATE-68
HUMAS2	Human argininosuccinate synthetase gene, exon 2 and flanks. 102BP	PRIMATE-68
HUMAS3	Human argininosuccinate synthetase gene, exon 3 and flanks. 150BP	PRIMATE-68
HUMAS3PS	Human argininosuccinate synthetase pseudogene 3. 1868BP	PRIMATE-69
HUMAS4	Human argininosuccinate synthetase gene, exon 4 and flanks. 109BP	PRIMATE-69
HUMAS5	Human argininosuccinate synthetase gene, exons 5, 6 and 7. 336BP	PRIMATE-69
HUMAS6	Human argininosuccinate synthetase gene, exon 8. 102BP	PRIMATE-70
HUMAS7	Human argininosuccinate synthetase gene, exon 9 and flanks. 131BP	PRIMATE-70
HUMAS7PS1	Human argininosuccinate synthetase pseudogene 7, segment 1. 263BP	PRIMATE-70
HUMAS7PS2	Human argininosuccinate synthetase pseudogene 7, segment 2. 297BP	PRIMATE-71
HUMAS7PS3	Human argininosuccinate synthetase pseudogene 7, segment 3. 178BP	PRIMATE-71
HUMAS8	Human argininosuccinate synthetase gene, exons 10, 11, 12 and 13. 808BP	PRIMATE-71
HUMASA	Human argininosuccinate synthetase mRNA, complete cds. 1547BP	PRIMATE-72
HUMASGPR1	Human asialoglycoprotein receptor H1 mRNA, complete cds. 1277BP	PRIMATE-72
HUMASGPR2	Human asialoglycoprotein receptor H2 mRNA, complete cds. 1309BP	PRIMATE-73
HUMAT31	Human antithrombin III gene, exon 1. 102BP	PRIMATE-73
HUMAT32	Human antithrombin III gene, exon 2. 432BP	PRIMATE-73
HUMAT33	Human antithrombin III gene, exon 3. 385BP	PRIMATE-74
HUMAT34	Human antithrombin III gene, exon 4. 452BP	PRIMATE-74
HUMAT35	Human antithrombin III gene, exon 5. 108BP	PRIMATE-75
HUMAT36	Human antithrombin III gene, exon 6. 267BP	PRIMATE-75



Entry Name	Description and Length	Page
HUMATIIIF	Human antithrombin III gene, polymorphism F. 853BP	PRIMATE-75
HUMATIIIS	Human antithrombin III gene, polymorphism S. 929BP	PRIMATE-76
HUMB2M	Human beta-2 microglobulin gene mRNA. 549BP	PRIMATE-77
HUMBHA	Human beta-hexosaminidase alpha chain mRNA, complete cds. 2397BP	PRIMATE-77
HUMBKVH01	Human homologue to bkV replication enhancer region: 5' end. 275BP	PRIMATE-77
HUMBKVH02	Human homologue to bkV replication enhancer region: 3' end. 549BP	PRIMATE-78
HUMBLUR14	Human non-alu 5' end of blurl4 clone. 260BP	PRIMATE-78
HUMBLYM1	Human Blym-1 transforming gene, complete coding region. 1004BP	PRIMATE-78
HUMC1A21	Human procollagen type I alpha-2 chain, AluI repeat in intron XII. 356BP	PRIMATE-78
HUMC1A22	Human procollagen type I alpha-2 chain, partial exon I. 1130BP	PRIMATE-79
HUMC1A23	Human procollagen type I alpha-2 chain, AluI repeat in 3' flank. 347BP	PRIMATE-79
HUMC1AIM1	Human (CRL 1262 variant allele) pro-alpha-1 type-I collagen gene, exons 31-27. 721BP	PRIMATE-79
HUMC1AIM2	Human (CRL 1262 variant allele) pro-alpha-1 type-I collagen gene, exon 26, partial. 56BP	PRIMATE-80
HUMC1AIN1	Human (CRL 1262; normal allele) pro-alpha-1 type-I collagen gene, exons 31-27. 1374BP	PRIMATE-80
HUMC1AIN2	Human (CRL 1262 normal allele) alpha-1 type-I collagen gene, exon 26. 56BP	PRIMATE-80
HUMC1PA11	Human collagen type I; pro-alpha-1 mRNA(part 1). 198BP	PRIMATE-81
HUMC1PA12	Human collagen type I; pro-alpha-1 mRNA(part 2). 123BP	PRIMATE-81
HUMC1PA13	Human collagen type I; pro-alpha-1 mRNA(part 3). 180BP	PRIMATE-81
HUMC1PA14	Human collagen type I; pro-alpha-1 mRNA(part 4). 133BP	PRIMATE-81
HUMC1PA2	Human collagen type I; pro-alpha-2 gene, exon 1. 251BP	PRIMATE-81
HUMC2A11	Human collagen type II; pro-alpha-1, exon 14 from 3' end. 102BP	PRIMATE-82
HUMC2A12	Human collagen type II pro-alpha-1, exon 4 from 3' end. 165BP	PRIMATE-82
HUMC3	Human complement component C3 mRNA, alpha and beta subunits, complete cds. 5067BP	PRIMATE-82
HUMC5	Human complement component C5 mRNA, clone J-16. 1703BP	PRIMATE-83
HUMC56IFI	Human interferon induced mRNA c56 coding a 56000 mr protein. 342BP	PRIMATE-84
HUMC9	Human complement component C9 mRNA, complete cds. 2461BP	PRIMATE-84
HUMC9M	Human complement component C9 mRNA. 2023BP	PRIMATE-84
HUMCABL	Human homologue (c-abl) of abelson murine leukemia virus. 192BP	PRIMATE-85
HUMCAL	Human calcitonin, mRNA, complete coding sequence. 791BP	PRIMATE-85
HUMCATF	Human fibroblast catalase gene, partial exon 1, complete exon 2. 1848BP	PRIMATE-86
HUMCG1A1	Human pro-alpha-1 type I collagen gene, exons 46-51. 543BP	PRIMATE-86
HUMCG1A2	Human procollagen type I alpha-2 chain, partial exon 1 mutation C-propeptide region. 93BP	PRIMATE-86
HUMCG1PA1	Human proalpha 1 (I) chain of type I procollagen mRNA (partial). 3347BP	PRIMATE-87
HUMCG1PA2	Human collagen typeI; pro-alpha-2 chain mRNA. 172BP	PRIMATE-87
HUMCG1PAT	Human collagen type I; pro-alpha-2 mRNA, 3' end. 2486BP	PRIMATE-88
HUMCG2A1	Human alpha-1(II) collagen gene COL2A1, partial cds. 4845BP	PRIMATE-88
HUMCG3A1M	Human alpha-1 type III collagen triple helical region and 3' mRNA. 1284BP	PRIMATE-89
HUMCG5B	Human chorionic gonadotropin (HCG) gene 5, beta-subunit. 1665BP	PRIMATE-90
HUMCG6B	Human chorionic gonadotropin (HCG) gene 6, beta-subunit. 1665BP	PRIMATE-90
HUMCGB	Human chorionic gonadotropin (hcg) beta subunit mRNA. 539BP	PRIMATE-91
HUMCGBA2	Human chorionic gonadotropin beta subunit (hcgb) a: exon2. 203BP	PRIMATE-91
HUMCGBA3	Human chorionic gonadotropin beta subunit (hcgb) a: exon3. 468BP	PRIMATE-91
HUMCGBB47	Human chorionic gonadotropin beta subunit mRNA, 5' flank, clone pCG-beta-474. 152BP	PRIMATE-91
HUMCGBB50	Human chorionic gonadotropin beta subunit mRNA, 5' flank, clone pCG-beta-507. 192BP	PRIMATE-92
HUMCGBBA1	Human chorionic gonadotropin beta subunit gene, exon 1, clone CG-beta-a. 79BP	PRIMATE-92
HUMCGBBA2	Human chorionic gonadotropin beta subunit gene, exon 2, clone CG-beta-a. 203BP	PRIMATE-92
HUMCGBBA3	Human chorionic gonadotropin beta subunit gene, exon 3, clone CG-beta-a. 468BP	PRIMATE-92
HUMCGBE1	Human chorionic gonadotropin beta subunit (hcgb) e left: exon3. 159BP	PRIMATE-93
HUMCGBE2	Human chorionic gonadotropin beta subunit (hcgb) e left: exon 2. 203BP	PRIMATE-93
HUMCGBE3	Human chorionic gonadotropin beta subunit (hcgb) e left: exon 1. 130BP	PRIMATE-93
HUMCGBE4	Human chorionic gonadotropin beta subunit (hcgb) e center: exon1. 107BP	PRIMATE-93
HUMCGBE5	Human chorionic gonadotropin beta subunit (hcgb) e center: exon2. 203BP	PRIMATE-93
HUMCGBE6	Human chorionic gonadotropin beta subunit (hcgb) e center: exon3. 475BP	PRIMATE-93
HUMCGBE7	Human chorionic gonadotropin beta subunit (hcgb) e right: exon1. 83BP	PRIMATE-93
HUMCGBEL1	Human chorionic gonadotropin beta subunit pseudogene, exon 3, clone CG-beta-e. 159BP	PRIMATE-94
HUMCGBEL2	Human chorionic gonadotropin beta subunit pseudogene, exon 2, clone CG-beta-e. 203BP	PRIMATE-94
HUMCGBEL3	Human chorionic gonadotropin beta subunit pseudogene, exon 1, clone CG-beta-e. 130BP	PRIMATE-94
HUMCGBEL4	Human chorionic gonadotropin beta subunit gene, exon 1, clone CG-beta-e. 107BP	PRIMATE-94
HUMCGBEL5	Human chorionic gonadotropin beta subunit gene, exon 2, clone CG-beta-e. 203BP	PRIMATE-95
HUMCGBEL6	Human chorionic gonadotropin beta subunit gene, exon 3, clone CG-beta-e. 475BP	PRIMATE-95
HUMCGBEL7	Human luteinizing hormone beta subunit gene, exon 1, clone pCG-beta-e. 83BP	PRIMATE-95
HUMCGBRA	Human chorionic gonadotropin beta subunit (hcgb) mRNA 5' ut. (from cdna clone 192BP	PRIMATE-95
HUMCGBRB	Human chorionic gonadotropin beta subunit (hcgb) mRNA 5' ut. (from cdna clone 152BP	PRIMATE-95
HUMCH15M	Human mRNA from chromosome 15 gene with homology to MHC-HLA-SB-1 intron A. 632BP	PRIMATE-96
HUMCHSBS	Human activated oncogene c-has/bas. 695BP	PRIMATE-96
HUMCMOS	Human humos gene homologous to transforming gene of mmsv. 1303BP	PRIMATE-96
HUMCRAS2P	Human c-Ha-ras2 oncogene, pseudogene. 1263BP	PRIMATE-97
HUMCRF	Human corticotropin-releasing factor (crf) gene. 2685BP	PRIMATE-97
HUMCRP	Human c-reactive protein (crp) (aa 107-138) mRNA. 96BP	PRIMATE-98
HUMCRYG01	Human gamma-crystallin-3 gene, exons 1 and 2. 472BP	PRIMATE-98
HUMCRYG02	Human gamma-crystallin-3 gene, exon 3. 399BP	PRIMATE-98
HUMCRYG03	Human gamma-crystallin-4 gene, exons 1 and 2. 674BP	PRIMATE-98
HUMCRYG04	Human gamma-crystallin-4 gene, exon 3. 432BP	PRIMATE-99
HUMCRYG05	Human gamma-crystallin-2 pseudogene, exons 1 and 2. 649BP	PRIMATE-99
HUMCRYG06	Human gamma-crystallin-2 pseudogene, exon 3. 432BP	PRIMATE-99
HUMCRYG11	Human gamma-crystallin-1 pseudogene, exons 1 and 2. 659BP	PRIMATE-100
HUMCRYG12	Human gamma-crystallin-1 pseudogene, exon 3. 431BP	PRIMATE-100
HUMCS1	Human chorionic somatomammotropin gene hCS-1, complete cds. 2301BP	PRIMATE-100