DNA Data Bank of Japan

DNA Database

Release 9, July 1991, including

1130 entries, 2002124 bases, 71561 lines

All the entries in this release have been collected,
reviewed, and entered by DNA Data Bank of Japan (DDBJ).
The entries will be incorporated also into EMBL and GenBank
databases, and an integrated database will be produced and
released. The entries made by DDBJ can be identified as such
in the integrated database by their accession numbers with
"D" at the beginning.

Prepared by:
Takashi Gojobori, Administrative Director
Yoshio Tateno, Executive Director
Naruya Saitou, Assistant Director
Hidenori Hayashida, Section Manager
Yoshihiro Ugawa, Section Manager
Yukiko Yamazaki, Information Reviewer
Motono Horie, Information Coordinator
Masako Iwase, Information Coordinator
Shigeko Suzuki, Information Distributor
Yuko Hasegawa, Information Distributor
Mary Shimoyama, International Collaborator
Yuriko Noguchi, Support Planner

    National Institute of Genetics
    DNA Research Center
    Laboratory of DNA Data Analysis
            Mishima 411, Japan
            Phone: +81 559 75 0771 x647
            FAX: +81 559 75 6040
            E-mail: ddbj@ddbj.nig.ac.jp

Files included:

1) ddbjrel.txt: release notes; this document
2) ddbjsub.txt: a data submission form
3) ddbj.dna: DNA data
4) ddbj.cds: peptide coding sequences extracted from ddbj.dna
5) ddbj.pep: peptide sequences translated from ddbj.cds
6) ddbj.dir: a short directory file
7) ddbj.acc: an accession number index file
8) ddbj.aut: an author name index file
9) ddbj.jou: a journal name index file

Files 4 (ddbj.cds) and 5 (ddbj.pep) in the above list were
generated by using "seqext" and "pepttr" programs,
respectively, written by Dr. Jim Fickett of GenBank, Los
Alamos National Laboratory.

Files 6-9 may include more than 80 characters in one line. If
this is the case, the line will be folded at every 81th column in
the file on the distribution tape with the fixed size of 80
bytes.

Acknowledgments:

We thank GenBank and EMBL Data Library for help. In particular
GenBank kindly provided us with tools that are useful for
quality control in data entry.

Appendix 1. DDBJ database release history

Release	Date	Entries	Bases	Lines
9	07/91	1130	2002124	71561
8	01/91	879	1573442	55622
7	07/90	681	1154211	41264
6	01/90	496	841236	29682
5	07/89	395	679378	23895
4	01/89	302	535985	18713
3	07/88	230	345850	13773
2	01/88	142	199392	8238
1	07/87	66	108970	4291

Appendix 2. An example of the DNA sequence file (ddbj.dna)

The format of this file is the same as one of GenBank data
which is described in the release note of GenBank Release 63.

```
------------------------------------------------------------------------
LOCUS       BOVAD2        325 bp ds-DNA              MAM       30-MAY-1990
DEFINITION  Bovine adrenodoxin gene, exon 2.
ACCESSION   D00469
KEYWORDS    Ad; adrenodoxin.
SEGMENT     2 of 4
SOURCE      Bovine adult liver, genomic DNA, clone lambdaGBAI.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Vertebrata; Tetrapoda; Mammalia;
            Eutheria; Artiodactyla.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Sagara,Y., Sawae,H., Kimura,A., Sagara-Nakano,Y., Morohashi,K.,
            Miyoshi,K. and Horiuchi,T.
  TITLE     Structural organization of the bovine adrenodoxin gene
  JOURNAL   J. Biochem. 107, 77-83 (1990)
  STANDARD  full staff_review
REFERENCE   2  (bases 86 to 245)
  AUTHORS   Kagimoto,M., Kagimoto,K., Simpson,E.R. and Waterman,M.R.
  TITLE     Transcription of the bovine adrenodoxin gene produces two species
            of mRNA of which only one is translated into adrenodoxin
  JOURNAL   J. Biol. Chem. 263, 8925-8928 (1988)
  STANDARD  full staff_review
REFERENCE   3  (bases 101 to 225)
  AUTHORS   Okamura,T., John,M.E., Zuber,M.X., Simpson,E.R. and Waterman,M.R.
  TITLE     Molecular cloning and amino acid sequence of the precursor form of
            bovine adrenodoxin: Evidence for a previously unidentified
            COOH-terminal peptide
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 82, 5705-5709 (1985)
  STANDARD  full staff_review
COMMENT     Submitted in computer readable form by Y. Sagara on 12-Sep-1989.
            The exon 3 in [2] correspond to exon 2 in [1].
FEATURES         from  to/span       description
     pept     +   101  +   225       adrenodoxin exon 2
     pre-msg  <     1  >   325       mRNA and introns
BASE COUNT      113 a     48 c     52 g     112 t
ORIGIN
        1 cacacacaac ctaattattt caccttacgg attcttcaag caactatttt aaagtatgtt
       61 gaaattactg aaagtaataa agaactgttt ccttttccag ctcagaagat aaaataacag
........
//
```

```
---------------------------------------------------------------------------

Appendix 3. An example of the peptide coding sequence file (ddbj.cds)

---------------------------------------------------------------------------
  ID          BOVAD2
              adrenodoxin exon 2
  LOCUS       BOVAD2          325 bp ds-DNA          MAM         30-MAY-1990
  ACCESSION   D00469
    ORGANISM  Bos taurus
              Eukaryota; Metazoa; Chordata; Vertebrata; Tetrapoda; Mammalia;
              Eutheria; Artiodactyla.
  FEATURES       from  to/span      description
       pept    +  101  +   225    0 adrenodoxin exon 2
  COMPLETE    5':n  3':n
  LENGTH      125
  ORIGIN
         1 ctcagaagat aaaataacag tccactttat aaaccgtgat ggtgaaacat taacaaccaa
        61 aggaaaaatt ggtgactctc tgctagatgt tgtggttcaa aataatctag atattgatgg
       121 ttttg
  //
  ...
  ------------------------------------------------------------------------


Appendix 4. An example of the peptide sequence file (ddbj.pep)

---------------------------------------------------------------------------
  ID          BOVAD2
              adrenodoxin exon 2
  LOCUS       BOVAD2          325 bp ds-DNA          MAM         30-MAY-1990
  ACCESSION   D00469
    ORGANISM  Bos taurus
              Eukaryota; Metazoa; Chordata; Vertebrata; Tetrapoda; Mammalia;
              Eutheria; Artiodactyla.
  FEATURES       from  to/span      description
       pept    +  101  +   225    0 adrenodoxin exon 2
  COMPLETE    5':n  3':n
  LENGTH      41
  ORIGIN      Translated using phase 2
         1 sedkitvhfi nrdgetlttk gkigdslldv vvqnnldidg f
  //
  ...
  ------------------------------------------------------------------------


Appendix 5. An example of the short directory file (ddbj.dir)

  Entry name #Acc            bp      Definition
  ---------------------------------------------------------------------------
  ABCADHCC    D00635 ds-DNA   4230 A. polyoxogenes alcohol dehydrogenase (EC 1.1.99.8) and cytochro
  ie c genes.
  ABCALDH     D00521 ds-DNA   2683 A.polyoxogenes membrane-bound aldehyde dehydrogenase gene, compl
  :te cds and flanks.
  ...................
  YSOGAPDH    D00134 ds-DNA   2159 Zygosaccharomyces rouxii, glyceraldehyde-3-phosphate dehydrogena
  :e gene (GAP-DH), complete cds.
  ZYMCP       D00593 ss-RNA   1963 Zucchini Yellow Mosaic Virus, nuclear inclusion protein and coat
  protein genes.
  ZYMV        D00692 ss-RNA   1546 Zucchini yellow mosaic virus gene, 3' terminal region, coat prot
  :in and RNA polymerase.
  ---------------------------------------------------------------------------


Appendix 6. An example of the accession number index file (ddbj.acc)
```

```
#Acc        Entry name #Primary accession
---------------------------------------------------
D00002 -> ECOPYRC     D00002
D00003 -> HUMP450M     D00003

.....................

X54488 -> CRECAH2G     D90207
Y00335 -> TOBPRIA      D90196
Y00335 -> TOBPRIBI     D90197
---------------------------------------------------
```

Appendix 7. An example of the author name index file (ddbj.aut)

```
Author     Entry name #Accession
-------------------------------------------------
Abad-Lapuebla,M.A. -> VIBTDHI     D90238
Abe,K. -> RICGLUTE    D00584
Abe,M. -> GOTGHRA     D00476

...

van Belkum,A. -> KYVCGA      D00637
van der Noordaa,J. -> PLYBCG      D00755
-------------------------------------------------
```

Appendix 8. An example of the journal name index file (ddbj.jou)

```
Journal name                          Entry name #Accession
-------------------------------------------------------------------
AIDS Res. Hum. Retrovirus 5, 593-604 (1989) -> HIV2GH1     M30895
Acta Obst. Gynaec. Jpn. 35, 1133-1142 (1983) -> MUSPRIMP   D00570
Actinomycetologica 5, 14-17 (1991) -> STMARGG      D00799

......................

Zool. Sci. 7, 879-887 (1990) -> RATPGHAA     D00575
Zool. Sci. 7, 879-887 (1990) -> RATTSHBA     D00578
-------------------------------------------------------------------
```

HELP NUC [DIR NUC]

The service provided by the EMBL Data Library allows users to download specific
entries and indices from the EMBL and GenBank Nucleotide Sequence databases.
Entries from the latest full release of the EMBL database are available, as
well as data created at EMBL and Genbank since that release.

Users can access a database entry as soon as it is created (at either EMBL or
GenBank), instead of having to wait for up to three months for it to appear
in the next release of the databases on magnetic tape or CD-ROM

The following files are available in the NUC directory:

    o   Every entry in the current release of the EMBL Nucleotide Sequence
        database, and any new entry created after that release was built.
        Each entry is stored in a separate file; the filename is the same as
        the entryname. Entries can be accessed either by ENTRY NAME or
        ACCESSION NUMBER.

    o   Entries created by GenBank since their latest release are also
        available and accessible by ACCESSION NUMBER. GenBank regularly
        transmits their newly-generated data to EMBL electronically for this
        purpose.

    o   The following indices built from entries new since the latest
        full releases of EMBL and GenBank databases.

            NEWSPECIES.NDX   - Species index
            NEWAUTHOR.NDX   - Author index
            NEWKEYWORD.NDX   - Keyword index
            NEWSHORTDIR.NDX   - Short directory index
            NEWCITATION.NDX   - Citation index
            NEWACNUMBER.NDX   - Accession number index

    o   NEWENTRIES.NDX   - A weekly listing of new EMBL and GenBank
                                       entries on the file server

An example:

A scientist reads about a sequence in a recent journal and wants to know if
it exists yet as a database entry. They decide to obtain the
citation index that includes only entries added to the database since the
latest full release, by sending the following command to the fileserver:

Get Nuc:NEWCITATION.NDX

They look up the reference, find that there is an entry called PIP03XX, and
obtain a copy of it by sending the following command to the file server:

Get Nuc:PIP03XX

If a database accession number is known, it can be used instead of the
entry name, and of course there is then no need to get the citation index.

The following files are available in the NUC directory:

o Every entry in the current release of the EMBL Nucleotide Sequence
   database, and any new or updated entry created after that release was
   built at EMBL, GenBank or DDBJ.
   Entries can be accessed either by ENTRY NAME or ACCESSION NUMBER.

   Current quarterly release of EMBL Nucleotide Sequence Database is:
   Release 27, May 1991, Freeze Date: 2.2.91

o The following indices built from entries new since latest full
   full release.

          NEWSPECIES.NDX   - Species index
          NEWAUTHOR.NDX   - Author index
          NEWKEYWORD.NDX   - Keyword index
          NEWSHORTDIR.NDX   - Short directory index
          NEWCITATION.NDX   - Citation index
          NEWACNUMBER.NDX   - Accession number index

o NEWENTRIES.NDX   - A weekly list of new and updated entries
                             from EMBL and GenBank.

o RELNOTES.DOC   - Release Notes for the latest release of
                             the EMBL Nucleotide Sequence Database

o USRMAN.DOC   - The EMBL Nucleotide Sequence Database
                             user manual which describes the format
                             and conventions used in the database.

2-Jul-1991