

GENBANK®



Subject Coverage	- Nucleic acid sequences and related descriptive data such as source organism, description, sequence length, and references - Contiguous Sequence (CONTIG) data			
File Type	Bibliographic, Sequence, Substance			
Features	Alerts (SDIs) CAS Registry Number® Identifiers Keep & Share Structures	Weekly ☒ ☒ ☐	Page Images SLART	☐ ☒
Record Content	- GenBank (registered trademark of the U.S. Department of Health and Human Services) is a nucleic acid database produced by the National Institute of Health. - Records in GenBank contain sequences and data such as the GenBank Locus Number, sequence description, source organism, sequence length, and references. - In addition, the file contains records with Contiguous Sequences (CONTIG) data consisting of a set of overlapping clones or sequences from which a sequence can be obtained. - Chemical Abstracts Service® has added CAS Registry Numbers (/RN) for nucleic acids and the CA Abstract Numbers (/OS) for the corresponding CA File references, thru 2016.			
File Size	More than 278 million records (2/2025)			
Coverage	1982 to the present			
Updates	- Updated daily - Reloaded bimonthly			
Language	English			
Database Producer	National Center for Biotechnology Information 8600 Rockville Pike Bethesda, MD 20892 USA			
Sources	The data are compiled primarily from journal literature and direct author submissions for otherwise unpublished sources			
User Aids	- Online Helps (HELP DIRECTORY lists all help messages available) - STNGUIDE			
Clusters	- AGRICULTURE - ALLBIB - AUTHORS - BIOSCIENCE - CASRNS - CHEMISTRY STN Database Clusters information (PDF).			
Pricing	Enter HELP COST at an arrow prompt (=>).			

Search and Display Field Codes

Nucleic acid sequences displayed in the SEQ field are not searchable in the GenBank File. For sequences added in 2016 and earlier, search the sequences in the REGISTRY File and then search the resulting REGISTRY answer set L-number in the GenBank File.

Field that allows left truncation (/BI) is marked with an asterisk (*).

Search Field Name	Search Code	Search Examples	Display Codes
Basic Index * (contains single words from the class identifier (CI), definition (DEF), feature (FEAT), GenBank accession number (GBN), locus (LOC), organism (ORGN), origin (SRT), supplementary term (ST), title (TI), and Comment (COMMENT) fields, as well as CAS Registry Numbers) (1)	None (or /BI)	S MINICIRCLE DNA S REGULATED (L) EXCISION S MAJOR (P) STRAIN S CHROMOSOME (S) CLONE S 139791-74-5 S ?PLAST?	CI, COMMENT, DEF, FEAT, GBN, LOC, ORGN, RN, SRT, ST, TI
Author	/AU	S COOK,G?/AU S COOK, G?/AU	AU
Class Identifier	/CI	S L1 AND BACTERIA/CI	CI
Contiguous Sequences (contains the GenBank Accession Numbers for the individual sequences in the contiguous sequence)	/CONT	S AC000098/CONT	CONT
Date (2)	/DATE	S DATE>=20000101 AND L8	DATE
Definition	/DEF	S T. CONGOLENSE/DEF	DEF
Entry Date (2)	/ED	S L1 AND ED>20000800	Not displayed
Feature	/FEAT	S REPEAT-REGION/FEAT	FEAT
Field Availability	/FA	S L1 AND ORGN/FA S L1 AND CONT/FA	Not displayed
Field Not Available	/FNA	S Y CHROMOSOME AND (286364-81-6 OR RN/FNA)	Not displayed
GenBank Accession Number (includes GenBank VERSION (VER))	/GBN	S M19751/GBN	GBN
Genome Project Identifier	/PJID	S 10729/PJID or S GENOMEPROJECT/PJID	PJID
Journal Title	/JT	S MOL. BIOCHEM. PARASIT?/JT	SO
Locus	/LOC	S TRBKPM2/LOC	LOC
Organism Name	/ORGN	S TRYPANOSOMA AVIUM/ORGN	ORGN
Other Source	/OS	S L2 AND CA/OS S 107:110309/OS	OS
Publication Year (2)	/PY	S 2000/PY	SO
Sequence Length (2)	/SQL	S 964/SQL S 500-1000/SQL S SQL>100000	SQL
Source (contains journal title, collation information, (volume, issue, pages), and publication year)	/SO	S (MOL BIOCHEM AND 109)/SO	SO
Supplementary Term (Keyword)	/ST (or /KW)	S OPERON FUSION/ST S CARBAMOYL/ST (S) SYNTHETASE/ST	ST
Update Date (2)	/UP	S L7 AND UP>20000831	Not displayed

(1) A term with left truncation must contain at least four characters.

(2) Numeric search field that may be searched using numeric operators or ranges.

(3) Combined numeric and text field. Numbers of bases are numeric and may be searched using numeric operators or ranges. Base terms are text terms.

DISPLAY and PRINT Formats

Any combination of formats may be used to display or print answers. Multiple codes must be separated by spaces or commas, e.g., D L1 1-5 TI AU. The fields are displayed in the order requested.

Hit-term highlighting is available in all fields except FEAT, SEQ, and VER. Highlighting must be on in order to use the HIT, KWIC, and OCC formats.

Format	Content	Examples
AU CI CONT DATE DEF FEAT GBN (1,2) LOC ORGN OS PJID RN SEQ SO SQL SRT ST (KW) TI (1)	Author Class Identifier (Molecule Type, Division Code) Contiguous Sequence Date Definition Feature (tabular display of Feature Key, Location, and Qualifier) GenBank Accession Number Locus Organism Name Other Source Genome Project Identifier CAS Registry Number Sequence Source Sequence Length Origin Supplementary Term Title	D L4 1-4 AU D CI 1,3-5 D CONT D DATE 5-10 D 1-3,7,8 DEF D FEAT D GBN 1-5 D L1 LOC 3 D 1,3,6 ORGN L5 D OS D PJID D RN 2 D L8 SEQ 1-3 D 1,4 SO D L1 SQL D SRT D ST L1 4 D TI 3,4 TOTAL
ALL SQIDE (IDE)	LOC, GBN, VER, RN, SQL, CI, DATE, DEF, ST, Source (ORGN), PJID, SRT, Comments, Reference (AU, TI, SO, OS), FEAT, SEQ, CONT (ALL is the default) Sequence Identifying Fields (LOC, GBN, VER, RN, SQL, PJID, DEF, FEAT, SEQ, CONT)	D ALL D IDE 3-5
HIT KWIC OCC (1)	Fields containing hit terms Hit terms with 20 words on either side (KeyWord In Context) Number of occurrences of hit terms and fields in which they occur	D L3 HIT D KWIC 2 NOH D OCC

(1) No online display fee for this format.

(2) For more information on the GenBank VERSION, see <http://www.ncbi.nlm.nih.gov/Sitemap/sequenceIDs.html>

SELECT, ANALYZE, and SORT Fields

The SELECT command is used to create E-numbers containing terms taken from the specified field in an answer set.

The ANALYZE command is used to create an L-number containing terms taken from the specified field in an answer set.

The SORT command is used to rearrange the search results in either alphabetic or numeric order of the specified field(s).

Field Name	Field Code	ANALYZE/ SELECT (1)	SORT
Author	AU	Y	N
CAS Registry Number	RN	Y (2)	Y
Class Identifier	CI	Y	N
Contiguous Sequence	CONT	Y (3)	N
Date	DATE	Y	Y
Definition	DEF	Y	Y
Feature	FEAT	Y (4)	N
GenBank Accession Number	GBN	Y (default)	Y
Genome Project Identifier	PJID	Y	Y
Journal Title	JT	Y (4)	N
Keyword	KW	Y	N
Locus	LOC	Y	Y
Occurrence of Hit Terms	OCC	N	Y
Organism Name	ORGN	Y	Y
Other Source	OS	Y	N
Publication Year	PY	Y (4)	N
Sequence Length	SQL	N	Y
Supplementary Term	ST	Y	N
Title	TI	Y	N

(1) HIT may be used to restrict terms extracted to terms that match the search expression used to create the answer set, e.g., SEL HIT RN.

(2) Appends /BI to the terms created by SELECT.

(3) Selects GBN and appends /CONT to the terms created by SELECT.

(4) SELECT HIT and ANALYZE HIT are not valid with this field.

Sample Records

DISPLAY ALL

LOCUS (LOC): BE185916 GenBank (R)
GenBank ACC. NO. (GBN): BE185916
GenBank VERSION (VER): BE185916.1
CAS REGISTRY NO. (RN): 274642-22-7
SEQUENCE LENGTH (SQL): 401
MOLECULE TYPE (CI): mRNA; linear
DIVISION CODE (CI): Expressed sequence tag
DATE (DATE): 10 May 2010
DEFINITION (DEF): IL5-HT0731-120500-088-d12 HT0731 Homo sapiens cDNA,
mRNA sequence.
KEYWORDS (ST): EST
SOURCE: Homo sapiens (human)
ORGANISM (ORGN): Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi; Mammalia; Eutheria; Euarchontoglires;
Primates; Haplorrhini; Catarrhini; Hominidae; Homo

COMMENT:

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=IL5-HT0731-120500-088-d12&t3=2000-05-12&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 361.

REFERENCE: 1 (bases 1 to 401)
AUTHOR (AU): Dias Neto,E.; Garcia Correa,R.; Verjovski-Almeida,S.;
Briones,M.R.; Nagai,M.A.; da Silva,W. Jr.; Zago,M.A.;
Bordin,S.; Costa,F.F.; Goldman,G.H.; Carvalho,A.F.;
Matsukuma,A.; Baia,G.S.; Simpson,D.H.; Brunstein,A.;
deOliveira,P.S.; Bucher,P.; Jongeneel,C.V.;
O'Hare,M.J.; Soares,F.; Brentani,R.R.; Reis,L.F.; de
Souza,S.J.; Simpson,A.J.
TITLE (TI): Shotgun sequencing of the human transcriptome with ORF
expressed sequence tags
JOURNAL (SO): Proc. Natl. Acad. Sci. U.S.A., 97 (7), 3491-3496 (2000)
OTHER SOURCE (OS): CA 132:261298

FEATURES (FEAT):

Feature Key	Location	Qualifier
source	1..401	/organism="Homo sapiens" /mol-type="mRNA" /db-xref="taxon:9606" /dev-stage="Adult" /clone-lib="HT0731" /note="Organ: head-neck; Vector: puc18; Site-1: SmaI; Site-2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed

under low stringency conditions."

DISPLAY ALL (cont'd)

SEQUENCE (SEQ):

```
1  cccctcccc  gggatttcaa  ggtccagcga  gagctcaccg  gacgccgccg  gaaccgcgac
61  gctttccaag  gcggggggccc  ctctcacggg  gcgaacccat  tccaggggcg  cctgcccttc
121  acaaagaaaa  gagaactctc  cccgggggctc  ccgccgggctt  ttccgggatc  ggtcgcgtta
181  ccgcactgga  cgcctcgcgg  ggccccatctc  cgccactccg  gattcgggga  tctgaacccg
241  actccctttc  gatcggccga  gggcaacgga  ggccatcgcc  cgtcccttcg  gaacggcgct
301  cgcccatctc  tcaggaccga  ctgacccatg  ttcaactgct  gttcacatgg  aacccttctc
361  cactgtggcc  ttcaaggtct  cgtttgaata  ttgtacta  c
```

DISPLAY ALL (CONTIG RECORD)

LOCUS (LOC): AE005172 GenBank (R)
GenBank ACC. NO. (GBN): AE005172
GenBank VERSION (VER): AE005172.1
SEQUENCE LENGTH (SQL): 14221815
MOLECULE TYPE (CI): DNA; linear
DIVISION CODE (CI): Contiguous sequences
DATE (DATE): 5 Mar 2010
DEFINITION (DEF): Arabidopsis thaliana chromosome 1 top arm, complete sequence.
SOURCE: Arabidopsis thaliana (thale cress)
ORGANISM (ORGN): Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
Tracheophyta; Spermatophyta; Magnoliophyta;
eudicotyledons; core eudicotyledons; rosids; malvids;
Brassicales; Brassicaceae; Arabidopsis
REFERENCE: 1 (bases 1 to 14221815)
AUTHOR (AU): Theologis,A.; Ecker,J.R.; Palm,C.J.; Federspiel,N.A.;
Kaul,S.; White,O.; Alonso,J.; Altaf,H.; Araujo,R.;
Bowman,C.L.; Brooks,S.Y.; Buehler,E.; Chan,A.; Chao,Q.;
Chen,H.; Cheuk,R.F.; Chin,C.W.; Chung,M.K.; Conn,L.;
Conway,A.B.; Conway,A.R.; Creasy,T.H.; Dewar,K.;
Dunn,P.; Etgu,P.; Feldblyum,T.V.; Feng,J.; Fong,B.;
Fujii,C.Y.; Gill,J.E.; Goldsmith,A.D.; Haas,B.;
Hansen,N.F.; Hughes,B.; Huizar,L.; Hunter,J.L.;
Jenkins,J.; Johnson-Hopson,C.; Khan,S.; Khaykin,E.;
Kim,C.J.; Koo,H.L.; Kremenetskaia,I.; Kurtz,D.B.;
Kwan,A.; Lam,B.; Langin-Hooper,S.; Lee,A.; Lee,J.M.;
Lenz,C.A.; Li,J.H.; Li,Y.; Lin,X.; Liu,S.X.; Liu,Z.A.;
Luros,J.S.; Maiti,R.; Marziali,A.; Militscher,J.;
Miranda,M.; Nguyen,M.; Nierman,W.C.; Osborne,B.I.;
Pai,G.; Peterson,J.; Pham,P.K.; Rizzo,M.; Rooney,T.;
Rowley,D.; Sakano,H.; Salzberg,S.L.; Schwartz,J.R.;
Shinn,P.; Southwick,A.M.; Sun,H.; Tallon,L.J.;
Tambunga,G.; Toriumi,M.J.; Town,C.D.; Utterback,T.; van
Aken,S.; Vaysberg,M.; Vysotskaia,V.S.; Walker,M.;
Wu,D.; Yu,G.; Fraser,C.M.; Venter,J.C.; Davis,R.W.
TITLE (TI): Sequence and analysis of chromosome 1 of the plant
Arabidopsis thaliana
JOURNAL (SO): Nature, 408 (6814), 816-820 (2000)
OTHER SOURCE (OS): CA 134:142622
REFERENCE: 2 (bases 1 to 14221815)
TITLE (TI): Direct Submission
JOURNAL (SO): Submitted (30-NOV-2000) USDA, Plant Gene Expression
Center, 800 Buchanan Street, Albany, CA 94710, USA

FEATURES (FEAT):

Feature Key	Location	Qualifier
source	1..14221815	/organism="Arabidopsis thaliana" /mol-type="genomic DNA" /db-xref="taxon:3702" /chromosome="1"

/map="top arm"

DISPLAY ALL (CONTIG RECORD) (cont'd)

CONTIG (CONT):

```
join(AC074298.1:1..298,AC007323.5:1..86293,  
AC023628.3:27500..103801,AC061957.3:1..65931,AC009273.2:1..76158,  
AC020622.3:184..47033,U89959.1:2161..55372,AC064879.3:1..91184,  
complement(AC022521.4:24628..121668),  
complement(AC009525.3:5320..99266),  
complement(AC006550.2:201..78341),complement(AC005278.1:49..71097),  
AC002560.2:153..100786,complement(AC003027.1:7674..119914),  
complement(AC002411.1:201..76670),  
complement(AC000104.1:202..107527),  
complement(AC002376.1:176..100079),  
complement(AC004809.1:201..89688),  
complement(AC005322.2:301..53958),  
complement(AC000098.1:7625..99300),AC005106.2:1..50549,  
AC007153.2:1..103023,AC009999.2:1..79494,AC024174.2:141..74116,  
AC025290.3:1..47000,AC068143.4:1..40700,  
complement(AC007592.4:6667..91310),  
complement(AC011001.2:5001..87752),  
complement(AC067971.5:301..111464),  
complement(AC022464.4:5570..102799),  
complement(AC007583.2:1127..106740),AC026875.4:1..98313,  
AC011438.3:337..86879,AC006932.8:9796..84046,  
AC003981.2:1131..132082,complement(AC000106.1:200..109000),  
complement(AC003114.1:201..59261),AC006416.2:1..16349,  
AC003970.1:1..94717,AC000132.1:1..142870,AC004122.1:1299..67566,  
AC005489.1:2613..100184,AC007067.4:912..80882,AC009398.6:1..45074,  
complement(AC007354.3:541..70116),  
complement(U95973.1:7053..113539),  
complement(AC007259.4:661..97146),AC011661.5:2004..84454,  
complement(AC007296.2:4..91566),  
complement(AC002131.1:30741..123189),  
complement(AC022522.2:1850..88643),AC025416.4:1..80226,  
AC025417.4:1244..104024,AC012187.3:501..74128,AC007357.2:1..96834,  
AC011810.8:2256..47626,AC027134.4:320..79579,  
AC027656.4:2895..78854,AC068197.4:1..33265,AC007576.3:1054..111971,  
AC012188.2:1..96128,AC010657.3:2522..69491,AC006917.6:1..128999,  
AC012189.3:1..23452,AC007591.2:1..113730,AC013453.1:1575..90801,  
AC034256.3:1..77321,AC010924.2:1..80370,AC006341.2:129..102073,  
complement(AC011808.4:26955..91593),  
complement(AC026237.4:45663..89231),  
complement(AC051629.3:36135..92359),  
complement(AC007651.2:201..111938),  
complement(AC026479.2:121..7604),  
complement(AC007843.6:69357..84894),AC022492.5:1..77194,  
AC034257.3:3214..98750,AC034106.4:1..74933,AC034107.5:1..15613,  
AC069551.4:918..74571,complement(AC013354.6:202..93434),  
complement(AC026238.2:202..44879),  
complement(AC011809.2:4124..108767),AC068602.2:616..85435,  
AC069143.3:1..58638,AC025808.8:2401..107871,  
complement(AC024609.2:59677..87774),  
complement(AC007797.7:2438..119942),  
complement(AC022472.2:19..91657),complement(AC026234.4:1..53707),  
complement(AC027665.4:54270..99031),AC069251.5:846..135378,  
complement(AC007369.2:201..102951),complement(AC012190.3:7..68986),  
complement(AC036104.3:201..55513),AC015447.8:1..98794,  
AC007727.2:1..83459,AC013482.2:1..78411,AC069252.4:1..91625,  
complement(AC073942.2:201..36030),complement(AC068562.4:1..40951),  
complement(AC006551.6:10458..107000),  
complement(AC003979.2:201..81654),  
complement(AF000657.1:6238..91949),  
complement(AC002311.1:26463..85855),AC005292.4:553..73552,  
AC007945.3:1088..33616,AC005990.1:3926..99722,AC002423.3:1..95168,  
AC002396.1:1..119878,AC000103.2:1393..86313,
```

complement(AC004133.3:14385..105863),AC079374.3:1..95393,
DISPLAY ALL (CONTIG RECORD) (cont'd)

complement(AC079281.2:713..97794),AC084221.5:9952..32028,
complement(AC079829.4:4202..95064),
complement(AC013427.3:201..86912),AC006535.7:1..91845,
AC005508.1:1..66454,AC000348.2:1..101482,AC004557.2:3733..99490,
AC005916.2:1..32413,AC012375.3:1..83614,AC079280.2:687..43073,
AC069471.2:333..109661,complement(AC021044.5:7373..121720),
complement(AC010155.3:10371..104163),
complement(AC007508.2:23480..100661),
complement(AC021043.4:5817..154716),AC068667.4:1..125800,
complement(AC079288.2:19523..35930),
complement(AC008030.4:16101..95204),AC022455.3:761..76832,
complement(AC074176.3:6187..72058),AC073506.9:1..54663,
complement(AC025295.5:561..81393),AC009917.2:7177..75599,
AC007060.3:1..96183,AC004135.1:1001..66731,
complement(AC000107.2:195..102393),AC004793.2:1..74699,
AC007654.2:4053..78347,AC027135.4:452..34231,
complement(AC074360.4:9773..107856),
complement(AC079041.4:23584..119420),
complement(AC074309.2:4457..75298),
complement(AC084165.1:5448..89759),AC084110.3:1..17502,
AC007767.3:1..125817,AC055769.2:3031..33100,AC017118.3:4218..90584,
AC006424.5:1..87844,AC021045.2:1121..58691,AC027035.3:2428..52789,
AC051630.2:1..88779,complement(AC069299.2:5628..34135),
complement(AC010164.2:26470..103443),
complement(AC022288.3:18024..80309),
complement(AC015446.4:10619..124578),
complement(AC007454.3:7849..88070),
complement(AC023913.5:32867..82214),
complement(AC023279.2:21379..117585),
complement(AC007894.2:38291..85653),
complement(AC018460.2:23820..110438),
complement(AC079605.11:3103..104807),AC069160.5:1..97088,
complement(AC023064.3:7412..63489),
complement(AC007887.9:201..158096),AC021198.2:1..76151,
AC027032.7:1..88767,complement(AC021666.3:35709..90659),
AC006228.5:1..104600,AC025781.6:7795..111630,AC079278.1:1..1225,
AC021199.5:1..68211,complement(AC007918.2:8315..115963),
AC025782.5:1..91841,AC079028.10:1..11182,AC068901.6:1..69032,
complement(AC020646.2:22118..103931),
complement(AC007505.4:1..134499))

In North America

CAS Customer Center:
P.O. Box 3012
Columbus, Ohio 43210-0012
U.S.A.

Phone: 800-753-4227 (North America)
614-447-3731 (worldwide)
E-mail: help@cas.org
Internet: www.cas.org

In Europe

CAS Customer Center EMEA
represented by
FIZ Karlsruhe - Leibniz-Institute for Information Infrastructure
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen
Germany

Phone: +49-721-9588 3155
E-mail: EMEAhelp@cas.org
Internet: www.fiz-karlsruhe.de

In Japan

JAICI
(Japan Association for International Chemical Information)
Nakai Building
6-25-4 Honkomagome, Bunkyo-ku
Tokyo 113-0021
Japan

Phone: +81-3-5978-3601 (Technical Service)
+81-3-5978-3621 (Customer Service)
E-mail: support@jaici.or.jp (Technical Service)
customer@jaici.or.jp (Customer Service)
Internet: www.jaici.or.jp