

Definition of AC, GI & version

- AC: It is a **stable** complete sequence record.
- GI & version number: It is a **unique** identifiers for the sequence data within a record.
- Any change occurs to the sequence data, the version number for that sequence is incremented by **one decimal** and a **new GI** number is assigned.

File sequence formats

- The files for DNA, RNA or protein sequences could be displayed in two formats:

Genbank (gb)

Fasta

- Remember we are speaking here about the type of sequence file format which applies for DNA, RNA and protein.
- But the nucleotide database is only for DNA and RNA

GenBank

Nucleotide

[Advanced](#)

[Help](#)

[Display Settings:](#) GenBank

[Send:](#)

Change region shown

Customize view

N.frontalis enolase gene

GenBank: X80474.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS X80474 1968 bp DNA linear PLN 14-NOV-2006
DEFINITION N.frontalis enolase gene.
ACCESSION X80474
VERSION X80474.1 GI:515826
KEYWORDS enolase.
SOURCE Neocallimastix frontalis
ORGANISM [Neocallimastix frontalis](#)
Eukaryota; Fungi; Neocallimastigomycota; Neocallimastigomycetes;
Neocallimastigales; Neocallimastigaceae; Neocallimastix.
REFERENCE 1
AUTHORS Durand,R., Fischer,M., Rascle,C. and Fevre,M.
TITLE Neocallimastix frontalis enolase gene, enol: first report of an
intron in an anaerobic fungus
JOURNAL Microbiology (Reading, Engl.) 141 (PT 6), 1301-1308 (1995)
PUBMED [7670633](#)
REFERENCE 2 (bases 1 to 1968)
AUTHORS Durand,R.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1994) R. Durand, University Lyon 1 CNRS UMR106,
Biologie Cellulaire Fongique, Bat. 405 69622, Villeurbanne Cedex,
FRANCE
FEATURES Location/Qualifiers
source 1..1968
/organism="Neocallimastix frontalis"
/mol_type="genomic DNA"

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

[Find in this Sequence](#)

Related information

[Protein](#)

[PubMed](#)

[Taxonomy](#)

Recent activity

[Turn Off](#) [Clear](#)

N.frontalis enolase gene Nucleotide

Clone DB: an integrated NCBI resource for clone-associated data

Clone DB: an integrated NCBI resource for clone-associated data. PubMed

GenBank file format provide information

- **Length** (example, 1400 bp in case of DNA/RNA. OR 102 aa in case of proteins)
- **Type** (i.e, DNA, cDNA, RNA, cRNA, linear, circular)
- **Accession number** (AC) (example, U49845)
- **GenIdentifier** (GI or gi) (example, 292659771)
- **Version** of accession number (example, U49845.3)
- **Organism** name
- **Download** fasta file
- **Not suitable** for use in programs and software.
- Each **10 bases** are separated.
- Each line of sequence has **60 bases**.
- Sequences must be **converted** from genbank to fasta format

Features in GenBank

Look for base positions

base positions for CDS

Protein sequences as fasta format.
Also, given the GI number, protein
name and ID

```
FEATURES             Location/Qualifiers
     source            1..1000
                        /organism="Oryza sativa"
                        /mol_type="mRNA"
                        /db_xref="taxon:4530"
                        /clone="OsHSP26-5"
                        /tissue_type="green leaf"
                        /clone_lib="lambda ZAP II"
     gene              1..1000
                        /gene="Oshsp26"
     CDS                107..826
                        /gene="Oshsp26"
                        /note="chloroplast-localized sm
                        /codon_start=1
                        /product="heat shock protein 26
                        /protein_id="BAA78385.1"
                        /db_xref="GI:4996284"
                        /translation="MAAPFALVSRVSPAARLI
AAQENRDNTAVDVHVNQDGGNQQGNAVQRRPI
LDTMDRMFDDVALGFATPRRSLATGEVRMP
VEDDALVIRGEHKKKEEGEGAEGSGDGWVKER
VLLVTVPKTEVERKVIDVQVQ"
     polyA_site        1000
                        /gene="Oshsp26"
                        /note="26 a nucleotides"
```

Each line of sequence in Genbank is 60 bp

Note that:
Each line is
displayed as
10 sequences
separated from
ach other

```
1  ttaaaatgaa  tccaaaccaa  aagataataa  ccattggttc  ggtctgtatg  acaattggaa
61  tggctaactt  aatattacaa  attggaaaca  taatctcaat  atggattagc  cactcaattc
121 aacttgggaa  tcaaaatcag  attgaaacat  gcaatcaaag  cgtcattact  tatgaaaaca
181 acacttgggt  aatcagaca  tatgttaaca  tcagcaacac  caactttgct  gctggacagt
241 cagtggtttc  cgtgaaatta  gcgggcaatt  cctctctctg  ccctgttagt  ggatgggcta
301 tatacagtaa  agacaacagt  gtaagaatcg  gttccaaggg  ggatgtgttt  gtcataaggg
361 aaccattcat  atcatgctcc  cccttggaat  gcagaacctt  cttcttgact  caaggggcct
421 tgctaaatga  caaacattcc  aatggaacca  ttaaagacag  gagcccatat  cgaaccctaa
481 tgagctgtcc  tattggtgaa  gttccctctc  catacaactc  aagatttgag  tcagtcgctt
541 ggtcagcaag  tgcttgatc  gatggcatca  attggctaac  aattggaatt  tctggcccag
601 acaatggggc  agtggctgtg  ttaaagtaca  acggcataat  aacagacact  atcaagagtt
661 ggagaaacaa  tatattgaga  acacaagagt  ctgaatgtgc  atgtgtaaat  ggttcttgct
721 ttactgtaat  gaccgatgga  ccaagtaatg  gacaggcctc  atacaagatc  ttcagaatag
781 aaaagggaaa  gatagtcaaa  tcagtcgaaa  tgaatgccc  taattatcac  tatgaggaat
841 gtcctgttta  tcctgattct  agtgaaatca  catgtgtgtg  cagggataac  tggcatggct
901 cgaatcgacc  gtgggtgtct  ttcaaccaga  atctggaata  tcagatagga  tacatatgca
961 gtgggatttt  cggagacaat  ccacgcccta  atgataagac  aggcagttgt  ggtccagtat
1021 cgtctaattg  agcaaatgga  gtaaaagggt  tttcattcaa  atacggcaat  ggtgtttgga
1081 tagggagaac  taaaagcatt  agttcaagaa  acggttttga  gatgatttgg  gatccgaacg
1141 gatggactgg  gacagacaat  aacttctcaa  taaagcaaga  tatcgtagga  ataaatgagt
1201 ggtcaggata  tagcgggagt  tttgttcagc  atccagaact  aacagggctg  gattgtataa
1261 gaccttgctt  ctgggttgaa  ctaatcagag  ggcgaccaa  agagaacaca  atctggacta
1321 gcgggagcag  catatccttt  tgtggtgtaa  acagtgacac  tgtgggttgg  tcttggccag
1381 acggtgctga  gttgccattt  accattgaca  agtaatt
```

Finding gene co-ordinate

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Help

Advanced

Display Settings: GenBank

Homo sapiens chromosome 6 genomic scaffold, GRCh38 Primary Assembly HSCR6_CTG1

NCBI Reference Sequence: NT_007592.16

[FASTA](#) [Graphics](#)

Change region shown

Send: Change region shown Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Related information

Assembly

BioProject

Component Of

Components (Core)

Full text in PMC

Gene

GeneView in dbSNP

HomoloGene

Identical GenBank Sequence

Map Viewer

Probe

Protein

PubMed

Go to:

LOCUS NT_007592 58393888 bp DNA linear CON 03-FEB-2014

DEFINITION Homo sapiens chromosome 6 genomic scaffold, GRCh38 Primary Assembly HSCR6_CTG1.

ACCESSION NT_007592 GPS_003205511 NT_007180 NT_007204 NT_007234 NT_007239 NT_007255 NT_007358 NT_007402 NT_007412 NT_007432 NT_007454 NT_019430 NT_023358 NT_023406 NT_023407 NT_023409 NT_023412 NT_023424 NT_023426 NT_023429 NT_023430 NT_023484 NT_023513 NT_026301 NT_026302 NT_027038 NT_027048 NT_027049 NT_027050 NT_027051 NT_027052 NT_027059 NT_028206 NT_028223 NT_029311 NT_029314 NT_029316 NT_029319 NT_031804 NT_033172 NT_033943 NT_034880 NT_034881 NT_078021 NW_004070866

VERSION NT_007592.16 GI:568815314

DBLINK BioProject: [PRJNA168](#)
Assembly: [GCF_000001405.26](#)

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 58393888)
CONSRM International Human Genome Sequencing Consortium
TITLE Finishing the euchromatic sequence of the human genome
JOURNAL Nature 431 (7011), 931-945 (2004)
PUBMED 15496913

Display Settings: GenBank

Send:

Homo sapiens chromosome 6 genomic scaffold, GRCh38 Primary Assembly HSCHR6_CTG1

NCBI Reference Sequence: NT_007592.16

[FASTA](#) [Graphics](#)

Go to:

LOCUS NT_007592 58393888 bp DNA 1
DEFINITION Homo sapiens chromosome 6 genomic scaffold, GRCh38 Primary Assembly
HSCHR6_CTG1.
ACCESSION NT_007592 GPS_003205511 NT_007180 NT_007204 NT_007234 NT_007239
NT_007255 NT_007358 NT_007402 NT_007412 NT_007432 NT_007454
NT_019430 NT_023358 NT_023406 NT_023407 NT_023409 NT_023412
NT_023424 NT_023426 NT_023429 NT_023430 NT_023484 NT_023513
NT_026301 NT_026302 NT_027038 NT_027048 NT_027049 NT_027050
NT_027051 NT_027052 NT_027059 NT_028206 NT_028223 NT_029311
NT_029314 NT_029316 NT_029319 NT_031804 NT_033172 NT_033943
NT_034880 NT_034881 NT_078021 NW_004070866
VERSION NT_007592.16 GI:568815314
DBLINK BioProject: [PRJNA168](#)
Assembly: [GCF_000001405.26](#)
KEYWORDS RefSeq.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 58393888)
CONSRMT International Human Genome Sequencing Consortium
TITLE Finishing the euchromatic sequence of the human genome

You can enter numbers to see
sequence in this region



Change region shown

Whole sequence (abbreviated view)
 Selected region

from: to:

Customize view

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

Related information

[Assembly](#)

[BioProject](#)

[Component Of](#)

[Components \(Core\)](#)

[Full text in PMC](#)

[Gene](#)

[GeneView in dbSNP](#)

[HomoloGene](#)

Amino Acids (aa) abbreviations

- A Alanine
- B Aspartic or Asparagine
- C Cysteine
- D Aspartic
- E Glutamic
- F Phenylalanine
- G Glycine
- H Histidine
- I Isoleucine
- J Lysine
- K Leucine
- L Methionine
- M
- N Asparagine
- O Pyrrolysine
- P Proline
- Q Glutamine
- R Arginine
- S Serine
- T Threonine
- U Selenocysteine
- V Valine
- W Tryptophan
- X Any
- Y Tyrosine
- Z Glutamic or Glutamine