Molecular data base

NUCLEOTIDE

The Entrez Nucleotide database is a collection of sequences from several sources, including:

- all the sequence data from GenBank, EMBL, and DDBJ, the members of the International Nucleotide Sequence Databases Collaboration (INSDC);
- NCBI-curated Reference Sequences (RefSeqs);
- submitted assemblies and annotations from the Third Party Annotation (TPA) database nucleotide;
- sequences extracted from structure records from the Protein Databank (PDB).

The number of bases in these databases is growing at an exponential rate.

GenBank

It is a comprehensive database that contains publicly available nucleotide sequences for more than 260,000 named organism, obtained primarily through:

a) submissions from individual laboratories
b) batch submissions from large-scale sequencing projects.

Remember that………..

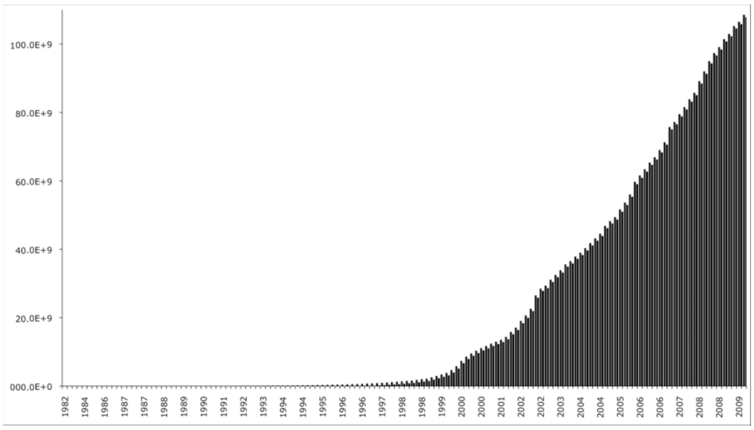
Data are shared nightly among the three collaborating databases

- GenBank at NCBI
- DNA Database of Japan (DDBJ)
- EMBL at EBI

International Nucleotide Sequence Database Collaboration
INSDC

NCBI-GenBank Release 180.0

118,551,641,086 bases
125,764,384 reported sequences



Number of entries and bases of DNA/RNA for the ten most sequenced organisms in Release 180.0,

excluding chloroplast and mitochondrial sequences, metagenomic sequences, Whole Genome Shotgun sequences, and 'constructed' CON-division sequences

ENTRIES	BASES	SPECIES
16403297	14891556303	Homo sapiens
7871531	8892650030	Mus musculus
2179595	6458987847	Rattus norvegicus
2190848	5362571338	Bos taurus
3892658	5037718386	Zea mays
3219186	4784078545	Sus scrofa
1698243	3125621603	Danio rerio
228238	1352920226	Strongylocentrotus purpuratus
1243345	1197439938	Oryza sativa Japonica Group
1756913	1187717551	Nicotiana tabacum
1424202	1147136093	Xenopus (Silurana) tropicalis
1212433	1049762304	Drosophila melanogaster
213953	1002175163	Pan troglodytes
2294285	1002146700	Arabidopsis thaliana
1452602	943091356	Canis lupus familiaris
660621	913979288	Vitis vinifera
810333	891564107	Gallus gallus
1880599	887191622	Glycine max
82244	823113612	Macaca mulatta
1216415	748356051	Ciona intestinalis

GenBank Divisions

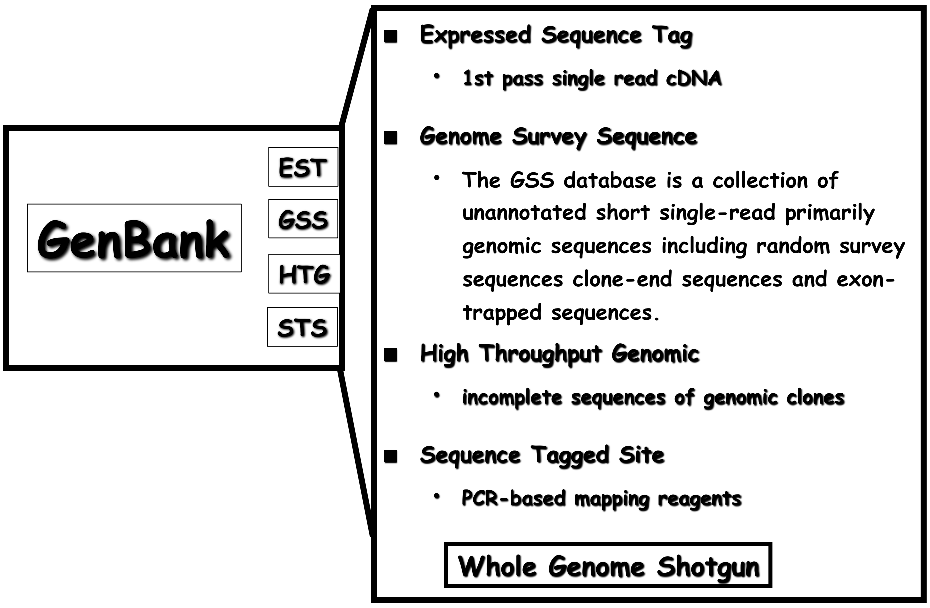
1. PRI - primate sequences
2. ROD - rodent sequences
3. MAM - other mammalian sequences
4. VRT - other vertebrate sequences
5. INV - invertebrate sequences
6. PLN - plant, fungal, and algal sequences
7. BCT - bacterial sequences
8. VRL - viral sequences
9. PHG - bacteriophage sequences
10. SYN - synthetic sequences
11. UNA - unannotated sequences

- Organized by taxonomy
- Direct submissions (Sequin/Bankit)
- Accurate (~1 error per 10,000 bp)
- Well characterized

GenBank Divisions

12. EST - EST sequences (expressed sequence tags)
13. PAT - patent sequences
14. STS - STS sequences (sequence tagged sites)
15. GSS - GSS sequences (genome survey sequences)
16. HTG - HTGS sequences (high throughput genomic sequences)
17. HTC - HTC sequences (high throughput cDNA sequences)
18. ENV - Environmental sampling sequences
19. CON - Constructed sequences

- Organized by sequence type
- Batch submissions (ftp/email)
- Less accurate
- Poorly characterized

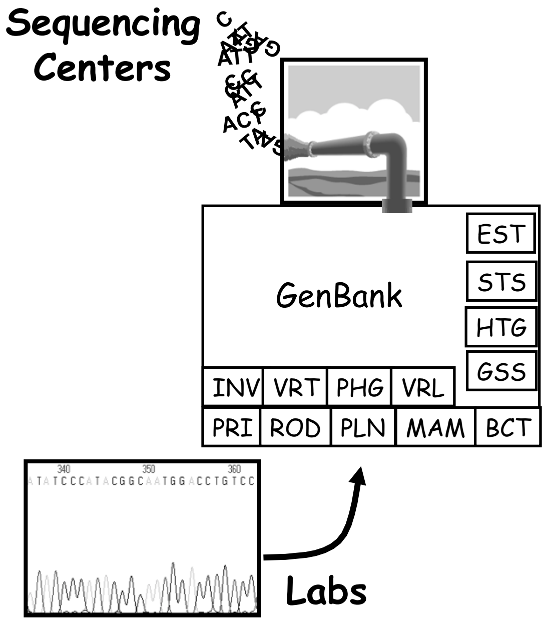


- **Expressed Sequence Tag**
  - **1st pass single read cDNA**
- **Genome Survey Sequence**
  - The GSS database is a collection of unannotated short single-read primarily genomic sequences including random survey sequences clone-end sequences and exon-trapped sequences.
- **High Throughput Genomic**
  - **incomplete sequences of genomic clones**
- **Sequence Tagged Site**
  - **PCR-based mapping reagents**

Whole Genome Shotgun

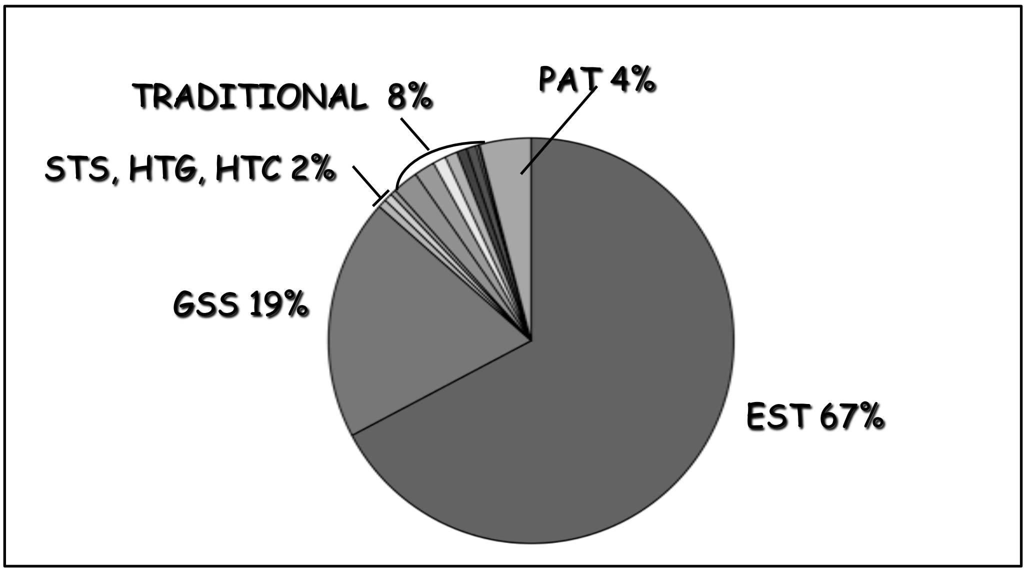
note

GENBANK IS A PRIMARY DATABASE



GenBank
Updated
ONLY
by submitters

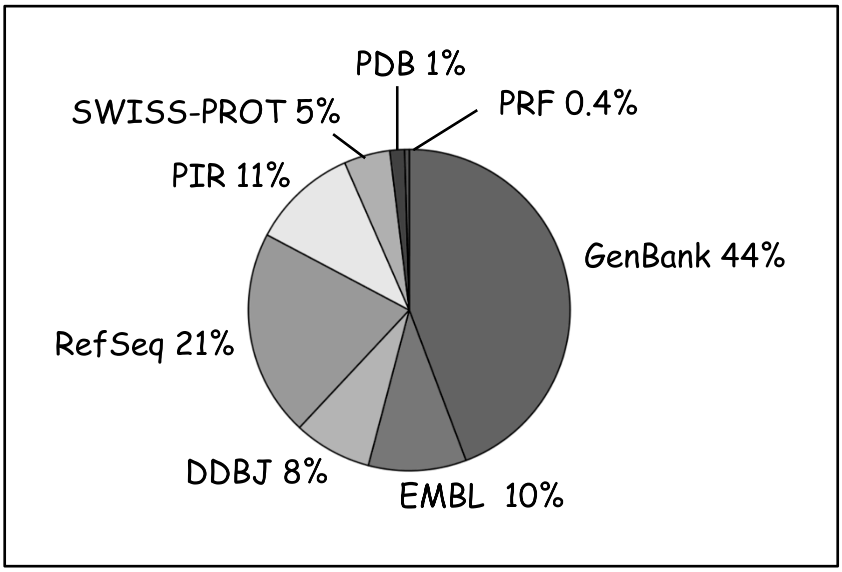
GenBank Organization



Protein Database

The protein entries in the Entrez search and retrieval system have been compiled from a variety of sources, including SwissProt, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq.

Protein database



Access to GenBank

There are several ways to search and retrieve data:

Search GenBank with Entrez Nucleotide

1. Keywords
2. Accession code

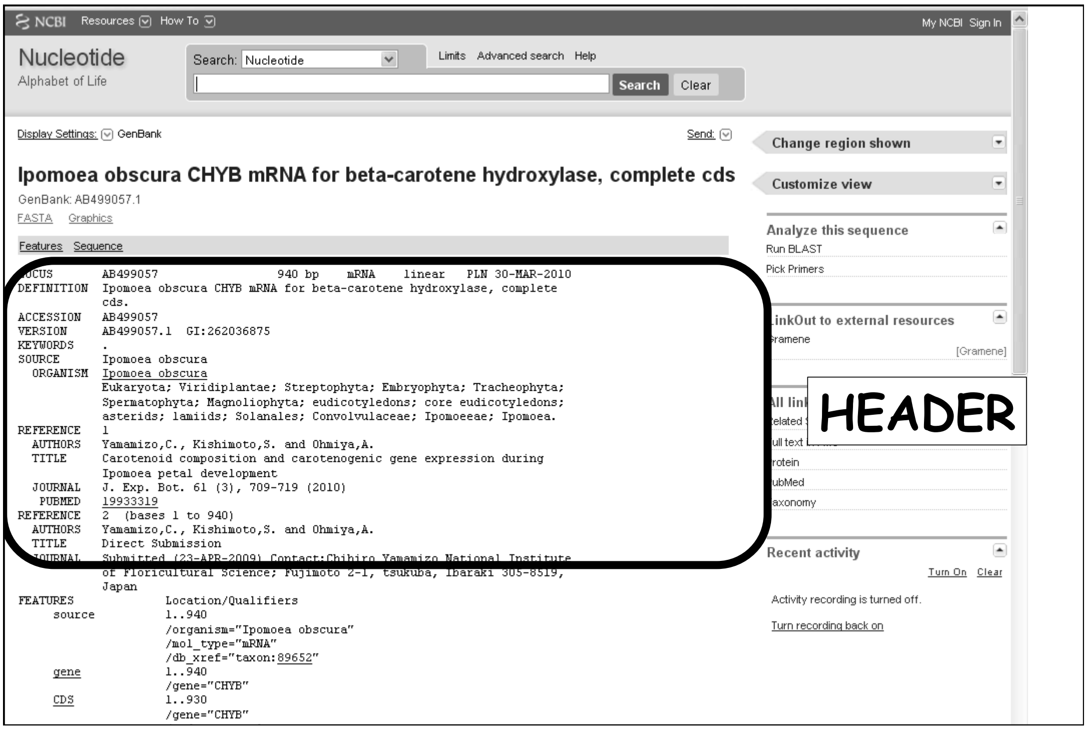
http://www.ncbi.nlm.nih.gov/nucleotide/

Beta carotene hydroxylase

link rete

Nucleotide - GenBank Record

- **Header**
  information that apply to the whole record
- **Features**
  annotations on the record
- **Sequence**



Ipomoea obscura CHYB mRNA for beta-carotene hydroxylase, complete cds

GenBank: AB499057.1

FASTA Graphics

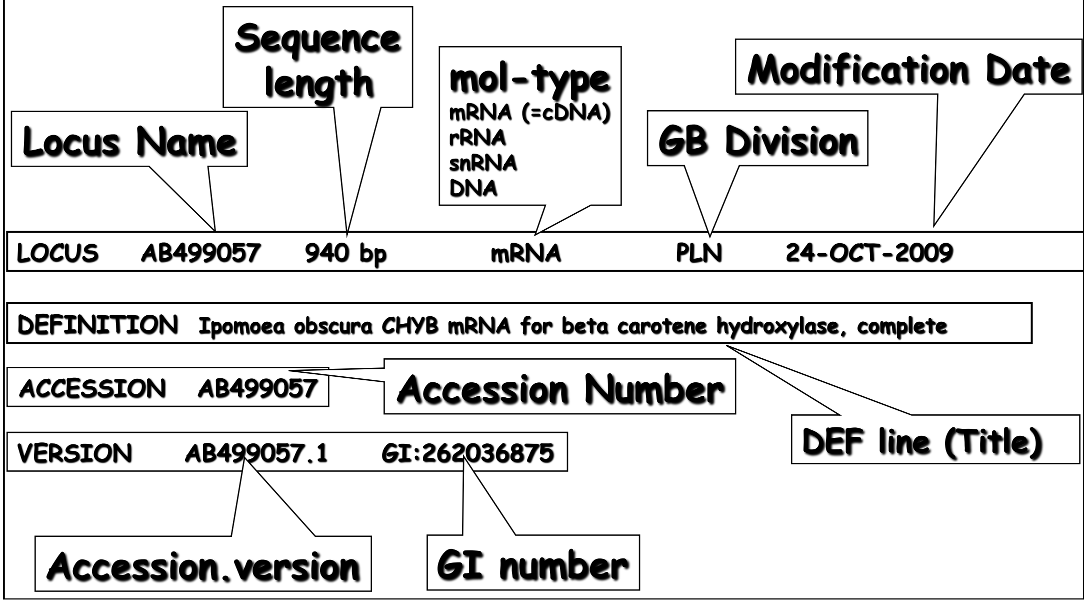
Features Sequence

```
OCUS       AB499057                 940 bp    mRNA    linear   PLN 30-MAR-2010
DEFINITION  Ipomoea obscura CHYB mRNA for beta-carotene hydroxylase, complete
            cds.
ACCESSION   AB499057
VERSION     AB499057.1  GI:262036875
KEYWORDS    .
SOURCE      Ipomoea obscura
  ORGANISM  Ipomoea obscura
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Convolvulaceae; Ipomoeeae; Ipomoea.
REFERENCE   1
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Carotenoid composition and carotenogenic gene expression during
            Ipomoea petal development
  JOURNAL   J. Exp. Bot. 61 (3), 709-719 (2010)
   PUBMED   19933319
REFERENCE   2  (bases 1 to 940)
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2009) Contact:Chihiro Yamamizo National Institute
            of Floricultural Science; Fujimoto 2-1, tsukuba, Ibaraki 305-8519,
            Japan
FEATURES             Location/Qualifiers
     source          1..940
                     /organism="Ipomoea obscura"
                     /mol_type="mRNA"
                     /db_xref="taxon:89652"
     gene            1..940
                     /gene="CHYB"
     CDS             1..930
                     /gene="CHYB"
```

HEADER: Citation

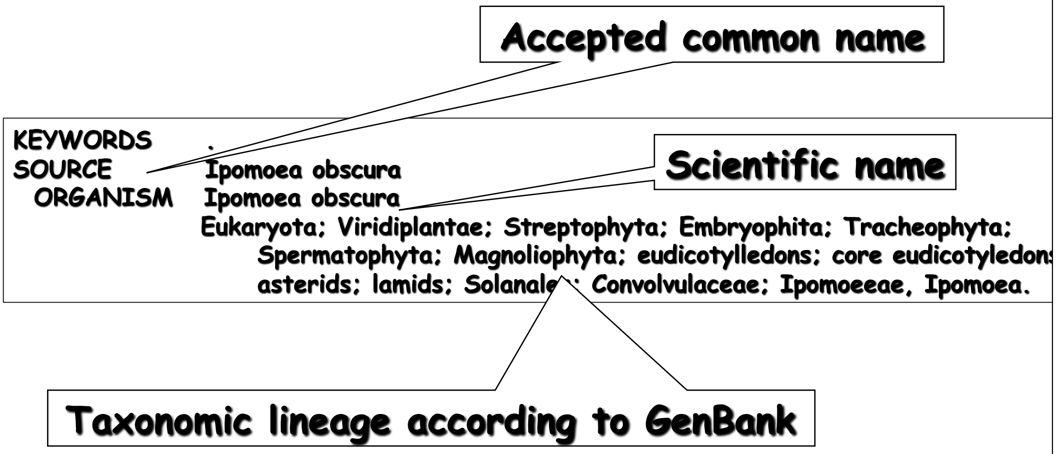
```
REFERENCE   1
  AUTHORS   Yamamizo, C., Kishimoto, S. and Ohmiya, A.
  TITLE     Analyses of carotenogenic gene expression during Ipomoea petal
            development
  JOURNAL   Unpublished
REFERENCE   2   (bases 1 to 940)
  AUTHORS   Yamamizo, C., Kishimoto, S. and Ohmiya, A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2009) Contact: Chihiro Yamamizo National Institute
             of Fluoricultural Science; Fujimoto 2-1, tsukuba, Ibaraki 305-8519,
             Japan
```

HEADER: database identifiers

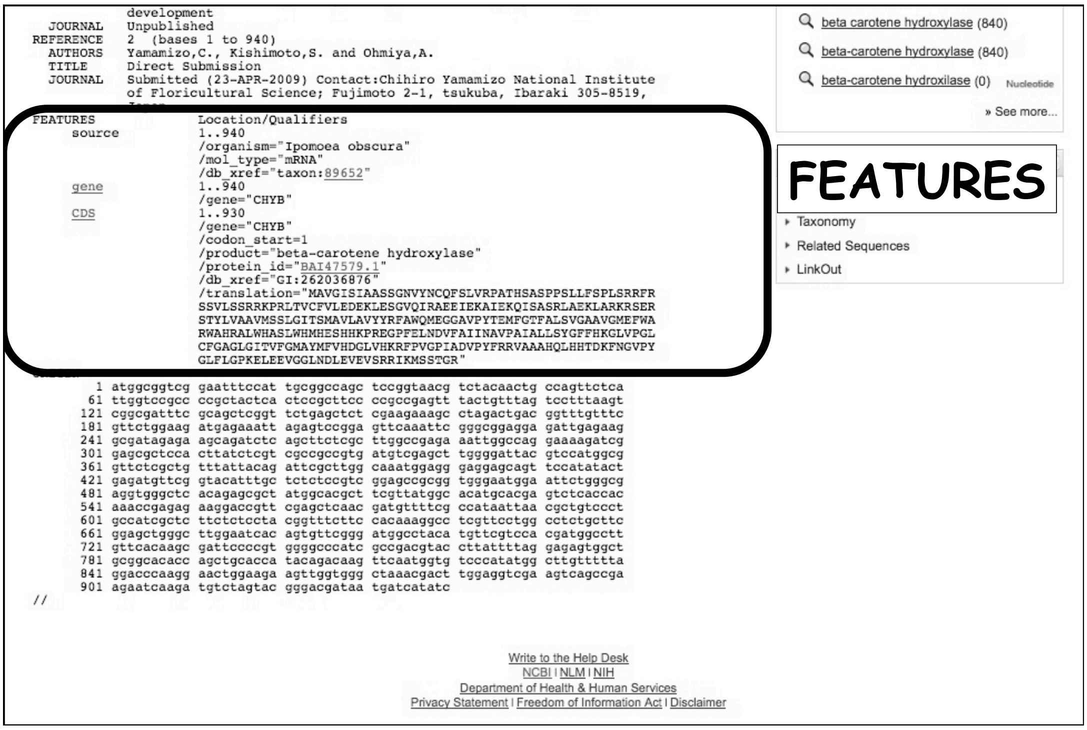


```
LOCUS     AB499057     940 bp          mRNA          PLN       24-OCT-2009

DEFINITION  Ipomoea obscura CHYB mRNA for beta carotene hydroxylase, complete

ACCESSION   AB499057

VERSION     AB499057.1     GI:262036875
```

HEADER: Keywords, Source, Organism



```
KEYWORDS        .
SOURCE          Ipomoea obscura
  ORGANISM      Ipomoea obscura
                Eukaryota; Viridiplantae; Streptophyta; Embryophita; Tracheophyta;
                    Spermatophyta; Magnoliophyta; eudicotylledons; core eudicotyledon
                    asterids; lamids; Solanales; Convolvulaceae; Ipomoeeae, Ipomoea.
```



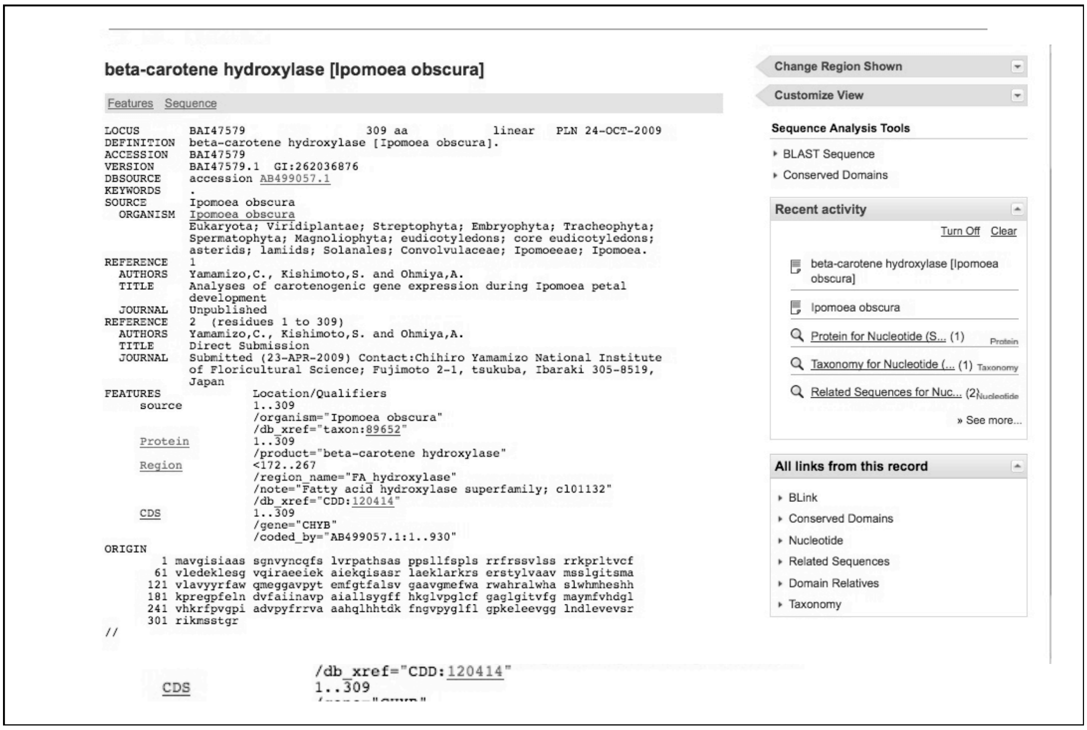
```
            development
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 940)
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2009) Contact:Chihiro Yamamizo National Institute
            of Floricultural Science; Fujimoto 2-1, tsukuba, Ibaraki 305-8519,
FEATURES             Location/Qualifiers
     source          1..940
                     /organism="Ipomoea obscura"
                     /mol_type="mRNA"
                     /db_xref="taxon:89652"
     gene            1..940
                     /gene="CHYB"
     CDS             1..930
                     /gene="CHYB"
                     /codon_start=1
                     /product="beta-carotene hydroxylase"
                     /protein_id="BAI47579.1"
                     /db_xref="GI:262036876"
                     /translation="MAVGISIAASSGNVYNCQFSLVRPATHSASPPSLLFSPLSRRFR
                     SSVLSSRRKPRLTVCFVLEDEKLESGVQIRAEEIEKAIEKQISASRLAEKLARKRSER
                     STYLVAAVMSSLGITSMAVLAVYYRFAWQMEGGAVPYTEMFGTFALSVGAAVGMEFWA
                     RWAHRALWHASLWHMHESHHKPREGPFELNDVFAIINAVPAIALLSYGFFHKGLVPGL
                     CFGAGLGITVFGMAYMFVHDGLVHKRFPVGPIADVPYFRRVAAAHQLHHTDKFNGVPY
                     GLFLGPKELEEVGGLNDLEVEVSRRIKMSSTGR"
        1 atggcggtcg gaatttccat tgcggccagc tccggtaacg tctacaactg ccagttctca
       61 ttggtccgcc ccgctactca ctccgcttcc ccgccgagtt tactgtttag tcctttaagt
      121 cggcgatttc gcagctcggt tctgagctct cgaagaaagc ctagactgac ggtttgtttc
      181 gttctggaag atgagaaatt agagtccgga gttcaaattc gggcggagga gattgagaag
      241 gcgatagaga agcagatctc agcttctcgc ttggccgaga aattggccag gaaaagatcg
      301 gagcgctcca cttatctcgt cgccgccgtg atgtcgagct tggggattac gtccatggcg
      361 gttctcgctg tttattacag attcgcttgg caaatggagg gaggagcagt tccatatact
      421 gagatgttcg gtacatttgc tctctccgtc ggagccgcgg tgggaatgga attctgggcg
      481 aggtgggctc acagagcgct atggcacgct tcgttatggc acatgcacga gtctcaccac
      541 aaaccgagag aaggaccgtt cgagctcaac gatgttttcg ccataattaa cgctgtccct
      601 gccatcgctc ttctctccta cggtttcttc cacaaaggcc tcgttcctgg cctctgcttc
      661 ggagctgggc ttggaatcac agtgttcggg atggcctaca tgttcgtcca cgatggcctt
      721 gttcacaagc gattccccgt ggggcccatc gccgacgtac cttattttag gagagtggct
      781 gcggcacacc agctgcacca tacagacaag ttcaatggtg tcccatatgg cttgttttta
      841 ggacccaagg aactggaaga agttggtggg ctaaacgact tggaggtcga agtcagccga
      901 agaatcaaga tgtctagtac gggacgataa tgatcatatc
//
```

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

Feature Table

```
FEATURES    Location/Qualifiers
   source      1..940
              /organism="Ipomoea obscura"
              /mol-type="mRNA"
              /db_xref="taxon:89652"
   CDS           1..932
              /gene="CHB1";
              /codon_start=1
              /product="beta-carotene hydroxylase"
              /protein_id="BAI47579.1"
              /db_xref="GI:262036876"
              /translation="MEYKCISEHLPFETLPDPGDRFEVQELVGTGTYATVYSAIDK
              NKKVALKIIGHIAENLLDIETEYRIYKAVNGIQFFPEFRGAFFKRGERESDNEVWL
"
```

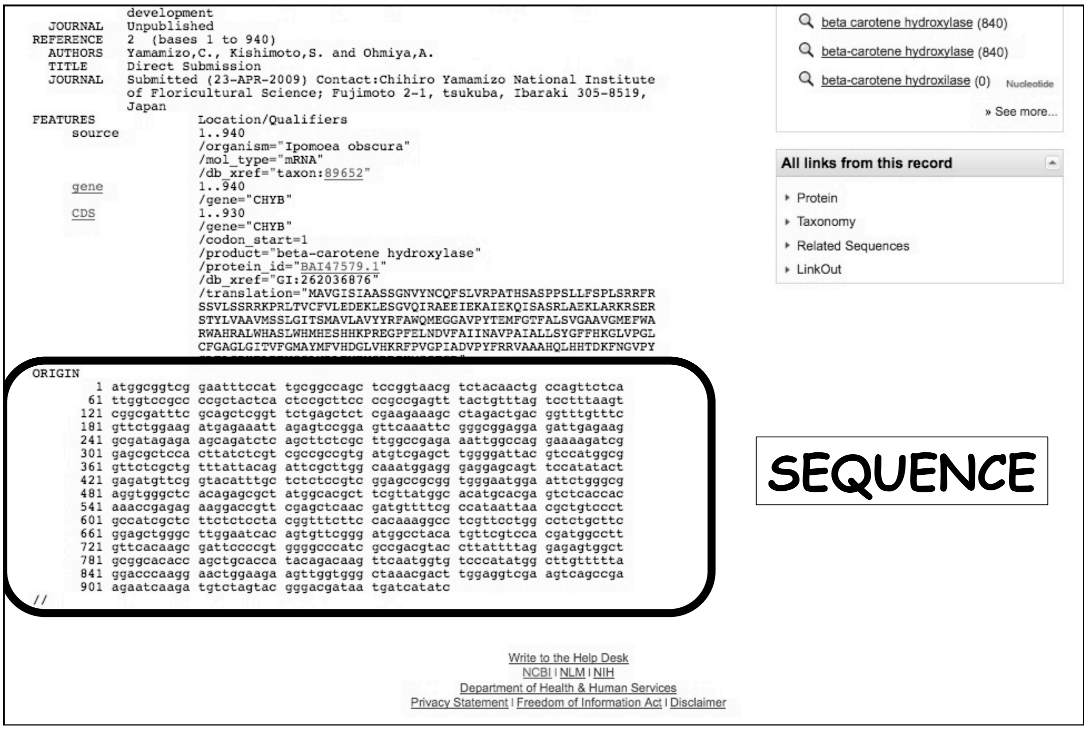
Direct link to protein records



beta-carotene hydroxylase [Ipomoea obscura]

```
LOCUS       BAI47579                 309 aa            linear   PLN 24-OCT-2009
DEFINITION  beta-carotene hydroxylase [Ipomoea obscura].
ACCESSION   BAI47579
VERSION     BAI47579.1  GI:262036876
DBSOURCE    accession AB499057.1
KEYWORDS    .
SOURCE      Ipomoea obscura
  ORGANISM  Ipomoea obscura
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Convolvulaceae; Ipomoeeae; Ipomoea.
REFERENCE   1
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Analyses of carotenogenic gene expression during Ipomoea petal
            development
  JOURNAL   Unpublished
REFERENCE   2  (residues 1 to 309)
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2009) Contact:Chihiro Yamamizo National Institute
            of Floricultural Science; Fujimoto 2-1, tsukuba, Ibaraki 305-8519,
            Japan
FEATURES             Location/Qualifiers
     source          1..309
                     /organism="Ipomoea obscura"
                     /db_xref="taxon:89652"
     Protein         1..309
                     /product="beta-carotene hydroxylase"
     Region          <172..267
                     /region_name="FA_hydroxylase"
                     /note="Fatty acid hydroxylase superfamily; cl01132"
                     /db_xref="CDD:120414"
     CDS             1..309
                     /gene="CHYB"
                     /coded_by="AB499057.1:1..930"
ORIGIN
        1 mavgisiaas sgnvyncqfs lvrpathsas ppsllfspls rrfrssvlss rrkprltvcf
       61 vledeklesg vqiraeeiek aiekqisasr laeklarkrs erstylvaav msslgitsma
      121 vlavyyrfaw qmeggavpyt emfgtfalsv gaavgmefwa rwahralwha slwhmheshh
      181 kpregpfeln dvfaiinavp aiallsygff hkglvpglcf gaglgitvfg maymfvhdgl
      241 vhkrfpvgpi advpyfrrva aahqlhhtdk fngvpyglfl gpkeleevgg lndlevevsr
      301 rikmsstgr
//
```

```
                     /db_xref="CDD:120414"
     CDS             1..309
```



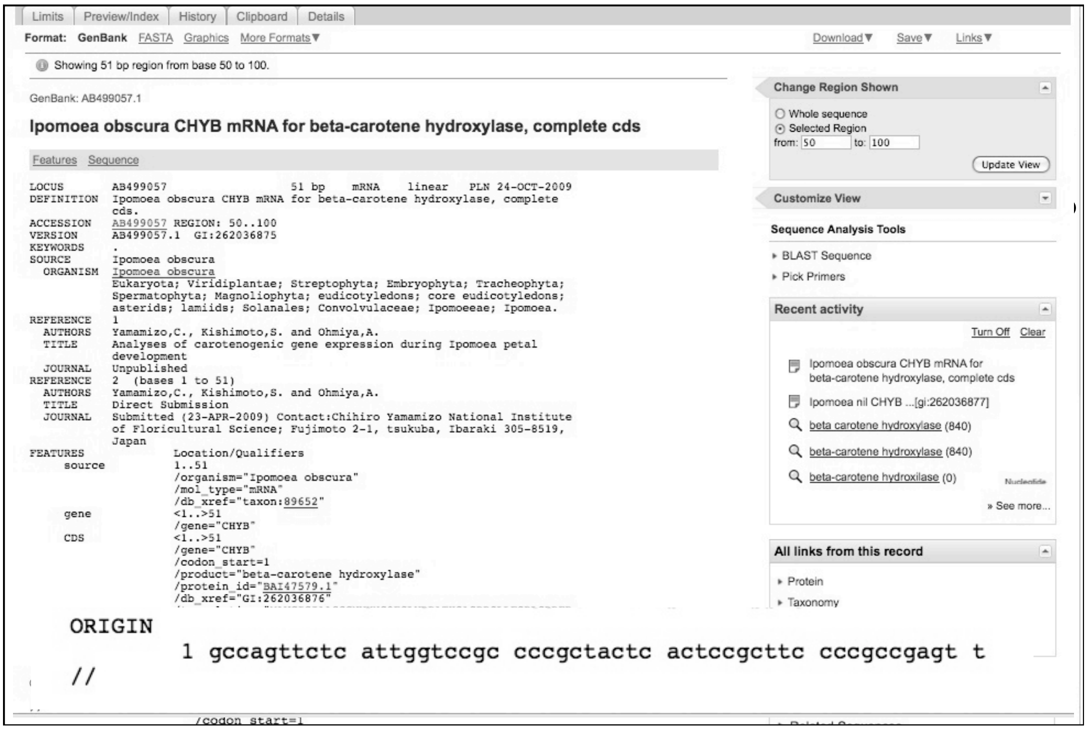
```
            development
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 940)
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2009) Contact:Chihiro Yamamizo National Institute
            of Floricultural Science; Fujimoto 2-1, tsukuba, Ibaraki 305-8519,
            Japan
FEATURES             Location/Qualifiers
     source          1..940
                     /organism="Ipomoea obscura"
                     /mol_type="mRNA"
                     /db_xref="taxon:89652"
     gene            1..940
                     /gene="CHYB"
     CDS             1..930
                     /gene="CHYB"
                     /codon_start=1
                     /product="beta-carotene hydroxylase"
                     /protein_id="BAI47579.1"
                     /db_xref="GI:262036876"
                     /translation="MAVGISIAASSGNVYNCQFSLVRPATHSASPPSLLFSPLSRRFR
                     SSVLSSRRKPRLTVCFVLEDEKLESGVQIRAEEIEKAIEKQISASRLAEKLARKRSER
                     STYLVAAVMSSLGITSMAVLAVYYRFAWQMEGGAVPYTEMFGTFALSVGAAVGMEFWA
                     RWAHRALWHASLWHMHESHHKPREGPFELNDVFAIINAVPAIALLSYGFFHKGLVPGL
                     CFGAGLGITVFGMAYMFVHDGLVHKRFPVGPIADVPYFRRVAAAHQLHHTDKFNGVPY
ORIGIN
        1 atggcggtcg gaatttccat tgcggccagc tccggtaacg tctacaactg ccagttctca
       61 ttggtccgcc ccgctactca ctccgcttcc ccgccgagtt tactgtttag tcctttaagt
      121 cggcgatttc gcagctcggt tctgagctct cgaagaaagc ctagactgac ggtttgtttc
      181 gttctggaag atgagaaatt agagtccgga gttcaaattc gggcggagga gattgagaag
      241 gcgatagaga agcagatctc agcttctcgc ttggccgaga aattggccag gaaaagatcg
      301 gagcgctcca cttatctcgt cgccgccgtg atgtcgagct tggggattac gtccatggcg
      361 gttctcgctg tttattacag attcgcttgg caaatggagg gaggagcagt tccatatact
      421 gagatgttcg gtacatttgc tctctccgtc ggagccgcgg tgggaatgga attctgggcg
      481 aggtgggctc acagagcgct atggcacgct tcgttatggc acatgcacga gtctcaccac
      541 aaaccgagag aaggaccgtt cgagctcaac gatgttttcg ccataattaa cgctgtccct
      601 gccatcgctc ttctctccta cggtttcttc cacaaaggcc tcgttcctgg cctctgcttc
      661 ggagctgggc ttggaatcac agtgttcggg atggcctaca tgttcgtcca cgatggcctt
      721 gttcacaagc gattccccgt ggggcccatc gccgacgtac cttattttag gagagtggct
      781 gcggcacacc agctgcacca tacagacaag ttcaatggtg tcccatatgg cttgttttta
      841 ggacccaagg aactggaaga agttggtggg ctaaacgact tggaggtcga agtcagccga
      901 agaatcaaga tgtctagtac gggacgataa tgatcatatc
//
```

SEQUENCE

Feature Table

```
FEATURES        Location/Qualifiers
    source        1..940
                 /organism="Ipomoea obscura"
                 /mol-type="mRNA"
                 /db_xref="taxon:89652"
    CDS            1..932
                /gene="CHB1";
                /codon_start=1
                /product="beta-carotene hydroxylase"
                /protein_id="BAI47579.1"
                /db_xref="GI:262036876"
                /translation="MEYKCISEHLPFETLPDPGDRFEVQELVGTGTYATVYSAIDK
                NKKVALKIIGHIAENLLDIETEYRIYKAVNGIQFFPEFRGAFFKRGERESDNEVWL
"
```

Coding sequence



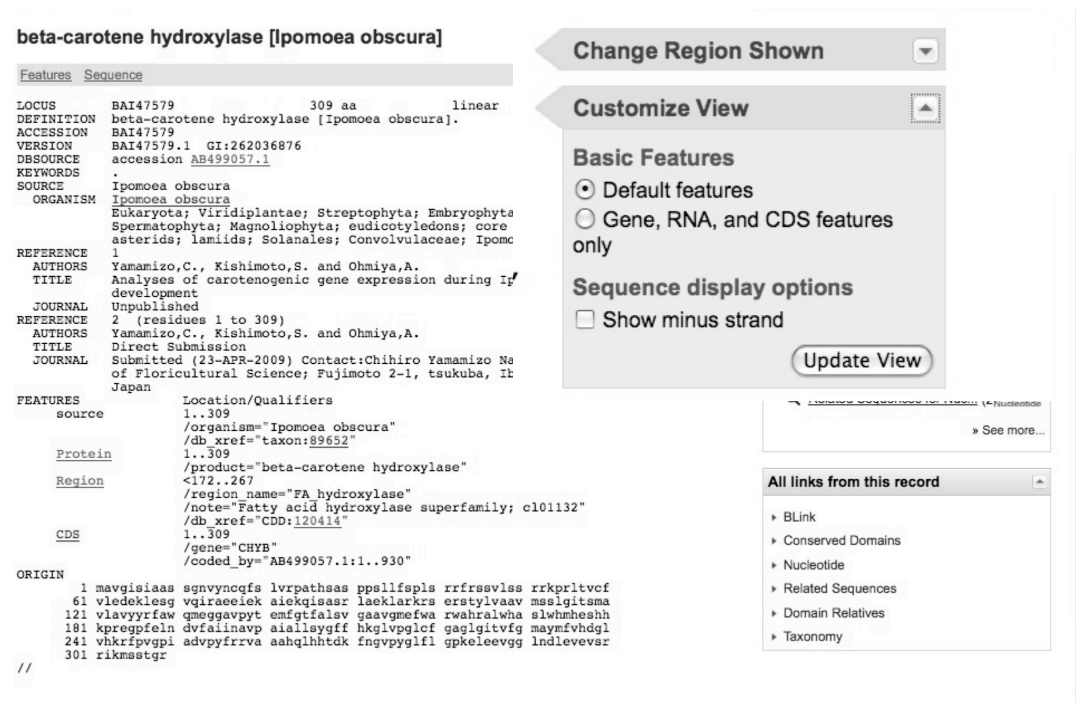
Showing 51 bp region from base 50 to 100.

from: 50 to: 100

GenBank: AB499057.1

Ipomoea obscura CHYB mRNA for beta-carotene hydroxylase, complete cds

```
LOCUS       AB499057                  51 bp    mRNA    linear   PLN 24-OCT-2009
DEFINITION  Ipomoea obscura CHYB mRNA for beta-carotene hydroxylase, complete
            cds.
ACCESSION   AB499057 REGION: 50..100
VERSION     AB499057.1  GI:262036875
KEYWORDS    .
SOURCE      Ipomoea obscura
  ORGANISM  Ipomoea obscura
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Convolvulaceae; Ipomoeeae; Ipomoea.
REFERENCE   1
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Analyses of carotenogenic gene expression during Ipomoea petal
            development
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 51)
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2009) Contact:Chihiro Yamamizo National Institute
            of Floricultural Science; Fujimoto 2-1, tsukuba, Ibaraki 305-8519,
            Japan
FEATURES             Location/Qualifiers
     source          1..51
                     /organism="Ipomoea obscura"
                     /mol_type="mRNA"
                     /db_xref="taxon:89652"
     gene            <1..>51
                     /gene="CHYB"
     CDS             <1..>51
                     /gene="CHYB"
                     /codon_start=1
                     /product="beta-carotene hydroxylase"
                     /protein_id="BAI47579.1"
                     /db_xref="GI:262036876"
ORIGIN
        1 gccagttctc attggtccgc cccgctactc actccgcttc cccgccgagt t
//
```



beta-carotene hydroxylase [Ipomoea obscura]

```
LOCUS       BAI47579                 309 aa            linear
DEFINITION  beta-carotene hydroxylase [Ipomoea obscura].
ACCESSION   BAI47579
VERSION     BAI47579.1  GI:262036876
DBSOURCE    accession AB499057.1
KEYWORDS    .
SOURCE      Ipomoea obscura
  ORGANISM  Ipomoea obscura
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta
            Spermatophyta; Magnoliophyta; eudicotyledons; core
            asterids; lamiids; Solanales; Convolvulaceae; Ipomc
REFERENCE   1
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Analyses of carotenogenic gene expression during Ip
            development
  JOURNAL   Unpublished
REFERENCE   2  (residues 1 to 309)
  AUTHORS   Yamamizo,C., Kishimoto,S. and Ohmiya,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2009) Contact:Chihiro Yamamizo Na
            of Floricultural Science; Fujimoto 2-1, tsukuba, Ib
            Japan
FEATURES             Location/Qualifiers
     source          1..309
                     /organism="Ipomoea obscura"
                     /db_xref="taxon:89652"
     Protein         1..309
                     /product="beta-carotene hydroxylase"
     Region          <172..267
                     /region_name="FA_hydroxylase"
                     /note="Fatty acid hydroxylase superfamily; cl01132"
                     /db_xref="CDD:120414"
     CDS             1..309
                     /gene="CHYB"
                     /coded_by="AB499057.1:1..930"
ORIGIN
        1 mavgisiaas sgnvyncqfs lvrpathsas ppsllfspls rrfrssvlss rrkprltvcf
       61 vledeklesg vqiraeeiek aiekqisasr laeklarkrs erstylvaav msslgitsma
      121 vlavyyrfaw qmeggavpyt emfgtfalsv gaavgmefwa rwahralwha slwhmheshh
      181 kpregpfeln dvfaiinavp aiallsygff hkglvpglcf gaglgitvfg maymfvhdgl
      241 vhkrfpvgpi advpyfrrva aahqlhhtdk fngvpyglfl gpkeleevgg lndlevevsr
      301 rikmsstgr
//
```

Basic Features

● Default features

○ Gene, RNA, and CDS features only

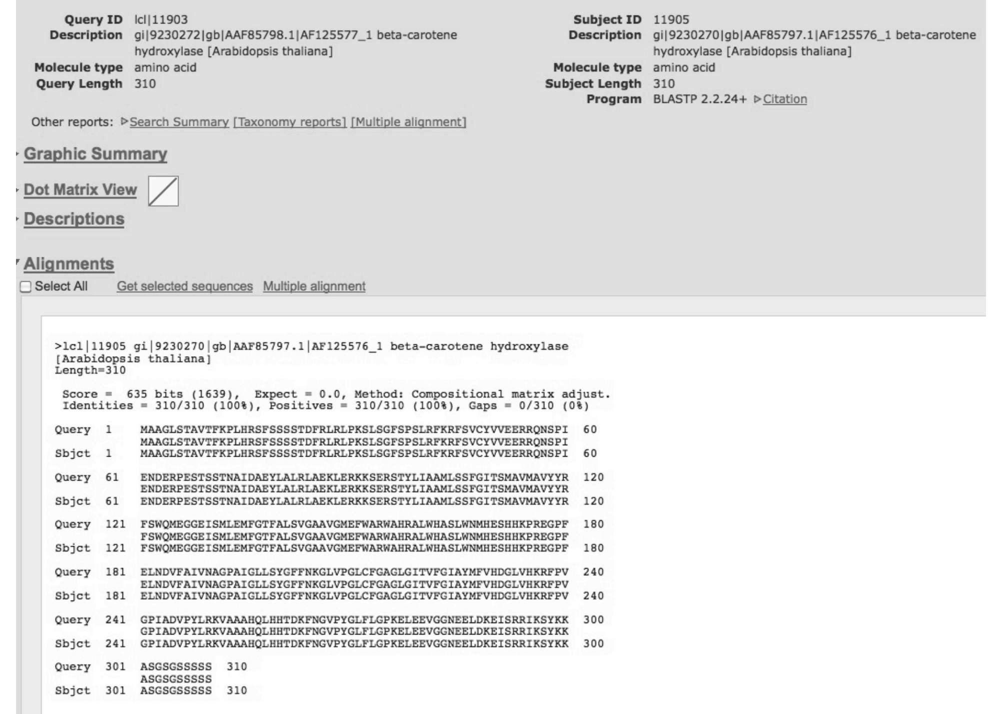
Sequence display options

☐ Show minus strand

GENBANK

Beta carotene hydroxylase
Arabidopsis thaliana

link rete



Query ID lcl|11903
Description gi|9230272|gb|AAF85798.1|AF125577_1 beta-carotene hydroxylase [Arabidopsis thaliana]
Molecule type amino acid
Query Length 310

Subject ID 11905
Description gi|9230270|gb|AAF85797.1|AF125576_1 beta-carotene hydroxylase [Arabidopsis thaliana]
Molecule type amino acid
Subject Length 310
Program BLASTP 2.2.24+ ▷Citation

```
>lcl|11905 gi|9230270|gb|AAF85797.1|AF125576_1 beta-carotene hydroxylase
[Arabidopsis thaliana]
Length=310

 Score =  635 bits (1639),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 310/310 (100%), Positives = 310/310 (100%), Gaps = 0/310 (0%)

Query  1    MAAGLSTAVTFKPLHRSFSSSSTDFRLRLPKSLSGFSPSLRFKRFSVCYVVEERRQNSPI  60
            MAAGLSTAVTFKPLHRSFSSSSTDFRLRLPKSLSGFSPSLRFKRFSVCYVVEERRQNSPI
Sbjct  1    MAAGLSTAVTFKPLHRSFSSSSTDFRLRLPKSLSGFSPSLRFKRFSVCYVVEERRQNSPI  60

Query  61   ENDERPESTSSTNAIDAEYLALRLAEKLERKKSERSTYLIAAMLSSFGITSMAVMAVYYR  120
            ENDERPESTSSTNAIDAEYLALRLAEKLERKKSERSTYLIAAMLSSFGITSMAVMAVYYR
Sbjct  61   ENDERPESTSSTNAIDAEYLALRLAEKLERKKSERSTYLIAAMLSSFGITSMAVMAVYYR  120

Query  121  FSWQMEGGEISMLEMFGTFALSVGAAVGMEFWARWAHRALWHASLWNMHESHHKPREGPF  180
            FSWQMEGGEISMLEMFGTFALSVGAAVGMEFWARWAHRALWHASLWNMHESHHKPREGPF
Sbjct  121  FSWQMEGGEISMLEMFGTFALSVGAAVGMEFWARWAHRALWHASLWNMHESHHKPREGPF  180

Query  181  ELNDVFAIVNAGPAIGLLSYGFFNKGLVPGLCFGAGLGITVFGIAYMFVHDGLVHKRFPV  240
            ELNDVFAIVNAGPAIGLLSYGFFNKGLVPGLCFGAGLGITVFGIAYMFVHDGLVHKRFPV
Sbjct  181  ELNDVFAIVNAGPAIGLLSYGFFNKGLVPGLCFGAGLGITVFGIAYMFVHDGLVHKRFPV  240

Query  241  GPIADVPYLRKVAAAHQLHHTDKFNGVPYGLFLGPKELEEVGGNEELDKEISRRIKSYKK  300
            GPIADVPYLRKVAAAHQLHHTDKFNGVPYGLFLGPKELEEVGGNEELDKEISRRIKSYKK
Sbjct  241  GPIADVPYLRKVAAAHQLHHTDKFNGVPYGLFLGPKELEEVGGNEELDKEISRRIKSYKK  300

Query  301  ASGSGSSSSS  310
            ASGSGSSSSS
Sbjct  301  ASGSGSSSSS  310
```

PROTEINS DATABASE

The Protein database contains amino acid sequences:

- created from the translations of coding regions provided on nucleotide records in GenBank, EMBL, and DDBJ
- those from coding regions on NCBI Reference Sequences and the Third Party Annotation (TPA) database records
- imported from the outside protein-only data sources Protein Information Resource (PIR), SWISS-PROT, Protein Research Foundation (PRF).
- extracted from structure records from the Protein Data Bank (PDB).

PROTEIN

Beta carotene hydroxylase
Ipomoea obscura

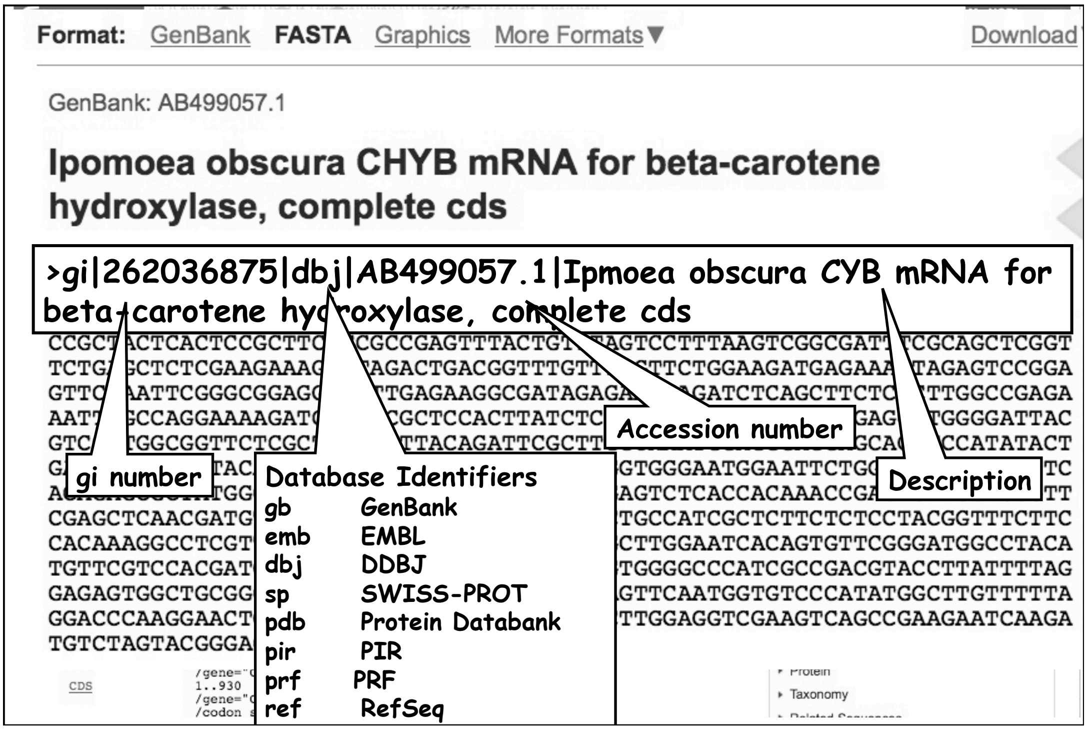
link rete

---

http://www.ncbi.nlm.nih.gov/protein/

link rete

FASTA FORMAT



GenBank: AB499057.1

Ipomoea obscura CHYB mRNA for beta-carotene hydroxylase, complete cds

```
>gi|262036875|dbj|AB499057.1|Ipmoea obscura CYB mRNA for beta-carotene hydroxylase, complete cds
```

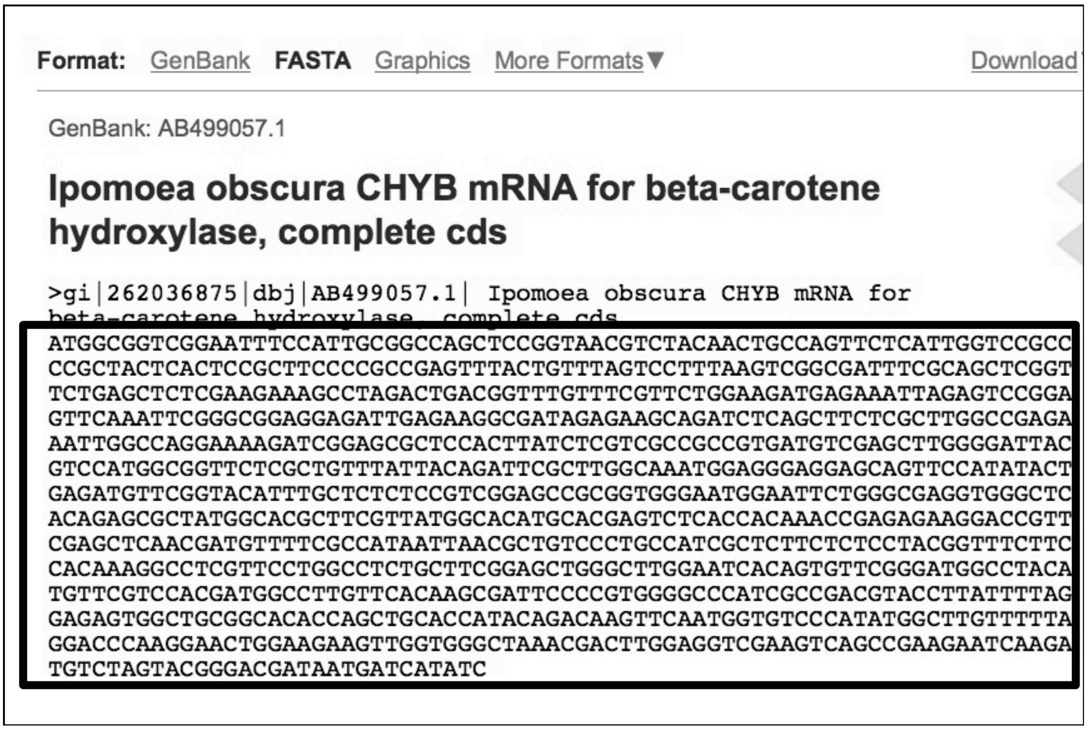
gi number

Accession number

Description

Database Identifiers

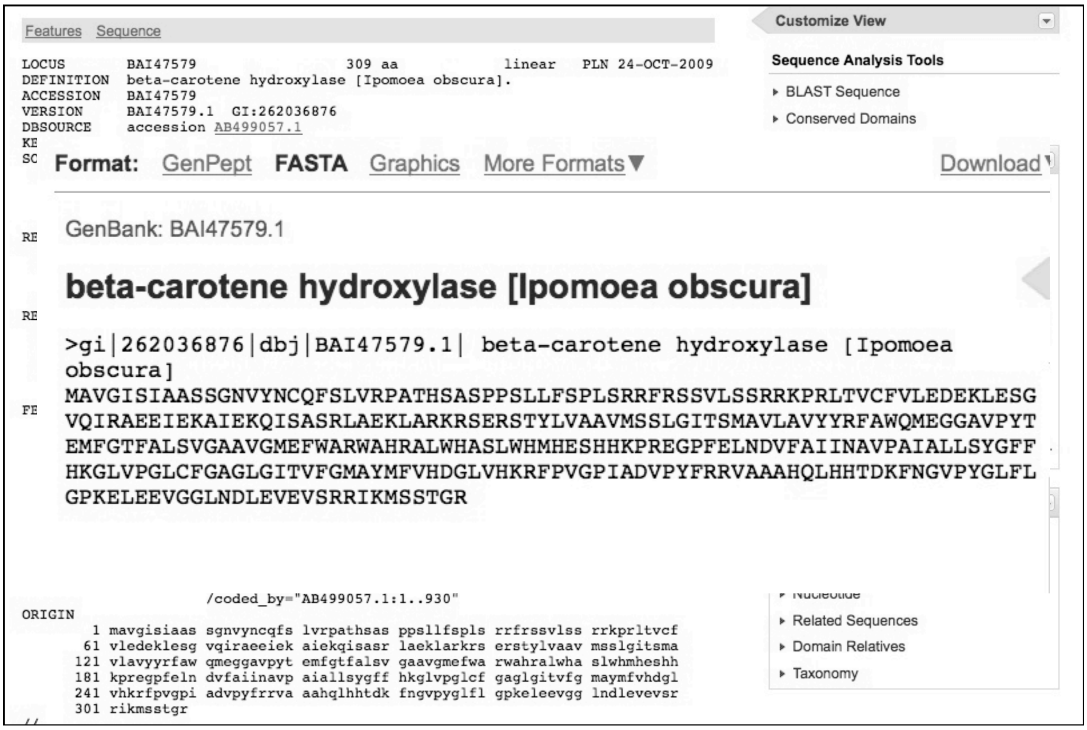
gb	GenBank
emb	EMBL
dbj	DDBJ
sp	SWISS-PROT
pdb	Protein Databank
pir	PIR
prf	PRF
ref	RefSeq



GenBank: AB499057.1

Ipomoea obscura CHYB mRNA for beta-carotene hydroxylase, complete cds

```
>gi|262036875|dbj|AB499057.1| Ipomoea obscura CHYB mRNA for
beta-carotene hydroxylase, complete cds
ATGGCGGTCGGAATTTCCATTGCGGCCAGCTCCGGTAACGTCTACAACTGCCAGTTCTCATTGGTCCGCC
CCGCTACTCACTCCGCTTCCCCGCCGAGTTTACTGTTTAGTCCTTTAAGTCGGCGATTTCGCAGCTCGGT
TCTGAGCTCTCGAAGAAAGCCTAGACTGACGGTTTGTTTCGTTCTGGAAGATGAGAAATTAGAGTCCGGA
GTTCAAATTCGGGCGGAGGAGATTGAGAAGGCGATAGAGAAGCAGATCTCAGCTTCTCGCTTGGCCGAGA
AATTGGCCAGGAAAAGATCGGAGCGCTCCACTTATCTCGTCGCCGCCGTGATGTCGAGCTTGGGGATTAC
GTCCATGGCGGTTCTCGCTGTTTATTACAGATTCGCTTGGCAAATGGAGGGAGGAGCAGTTCCATATACT
GAGATGTTCGGTACATTTGCTCTCTCCGTCGGAGCCGCGGTGGGAATGGAATTCTGGGCGAGGTGGGCTC
ACAGAGCGCTATGGCACGCTTCGTTATGGCACATGCACGAGTCTCACCACAAACCGAGAGAAGGACCGTT
CGAGCTCAACGATGTTTTCGCCATAATTAACGCTGTCCCTGCCATCGCTCTTCTCTCCTACGGTTTCTTC
CACAAAGGCCTCGTTCCTGGCCTCTGCTTCGGAGCTGGGCTTGGAATCACAGTGTTCGGGATGGCCTACA
TGTTCGTCCACGATGGCCTTGTTCACAAGCGATTCCCCGTGGGGCCCATCGCCGACGTACCTTATTTTAG
GAGAGTGGCTGCGGCACACCAGCTGCACCATACAGACAAGTTCAATGGTGTCCCATATGGCTTGTTTTTA
GGACCCAAGGAACTGGAAGAAGTTGGTGGGCTAAACGACTTGGAGGTCGAAGTCAGCCGAAGAATCAAGA
TGTCTAGTACGGGACGATAATGATCATATC
```



```
LOCUS       BAI47579                 309 aa            linear   PLN 24-OCT-2009
DEFINITION  beta-carotene hydroxylase [Ipomoea obscura].
ACCESSION   BAI47579
VERSION     BAI47579.1  GI:262036876
DBSOURCE    accession AB499057.1
```

GenBank: BAI47579.1

beta-carotene hydroxylase [Ipomoea obscura]

```
>gi|262036876|dbj|BAI47579.1| beta-carotene hydroxylase [Ipomoea
obscura]
MAVGISIAASSGNVYNCQFSLVRPATHSASPPSLLFSPLSRRFRSSVLSSRRKPRLTVCFVLEDEKLESG
VQIRAEEIEKAIEKQISASRLAEKLARKRSERSTYLVAAVMSSLGITSMAVLAVYYRFAWQMEGGAVPYT
EMFGTFALSVGAAVGMEFWARWAHRALWHASLWHMHESHHKPREGPFELNDVFAIINAVPAIALLSYGFF
HKGLVPGLCFGAGLGITVFGMAYMFVHDGLVHKRFPVGPIADVPYFRRVAAAHQLHHTDKFNGVPYGLFL
GPKELEEVGGLNDLEVEVSRRIKMSSTGR
```

```
                     /coded_by="AB499057.1:1..930"
ORIGIN
        1 mavgisiaas sgnvyncqfs lvrpathsas ppsllfspls rrfrssvlss rrkprltvcf
       61 vledeklesg vqiraeeiek aiekqisasr laeklarkrs erstylvaav msslgitsma
      121 vlavyyrfaw qmeggavpyt emfgtfalsv gaavgmefwa rwahralwha slwhmheshh
      181 kpregpfeln dvfaiinavp aiallsygff hkglvpglcf gaglgitvfg maymfvhdgl
      241 vhkrfpvgpi advpyfrrva aahqlhhtdk fngvpyglfl gpkeleevgg lndlevevsr
      301 rikmsstgr
```

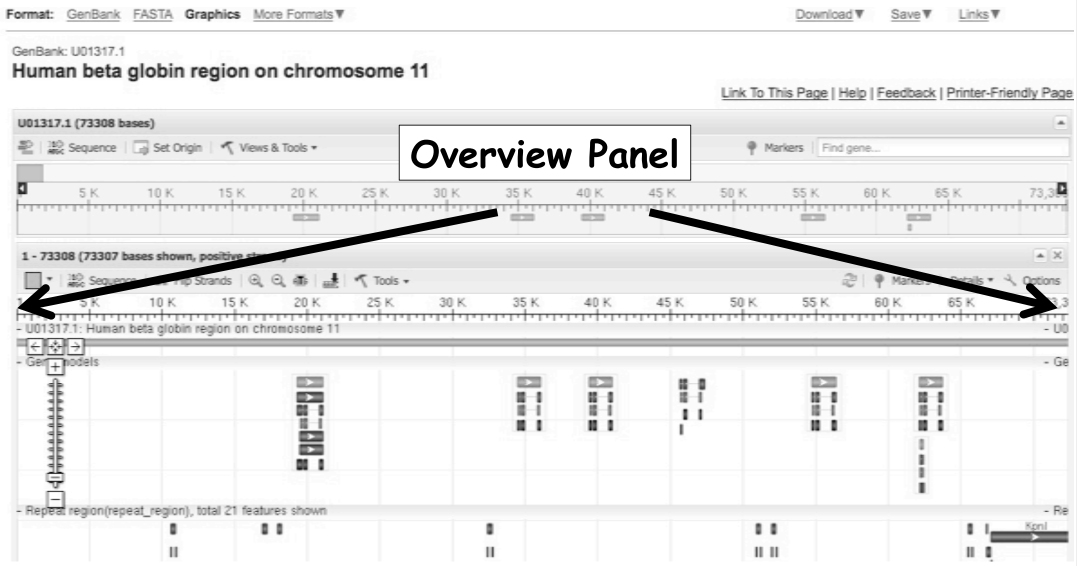
Nucleotide and protein graphical panel

The overview panel provides a brief glimpse of the entire sequence.

Human beta globin
AND chromosome 11
AND human

http://www.ncbi.nlm.nih.gov/

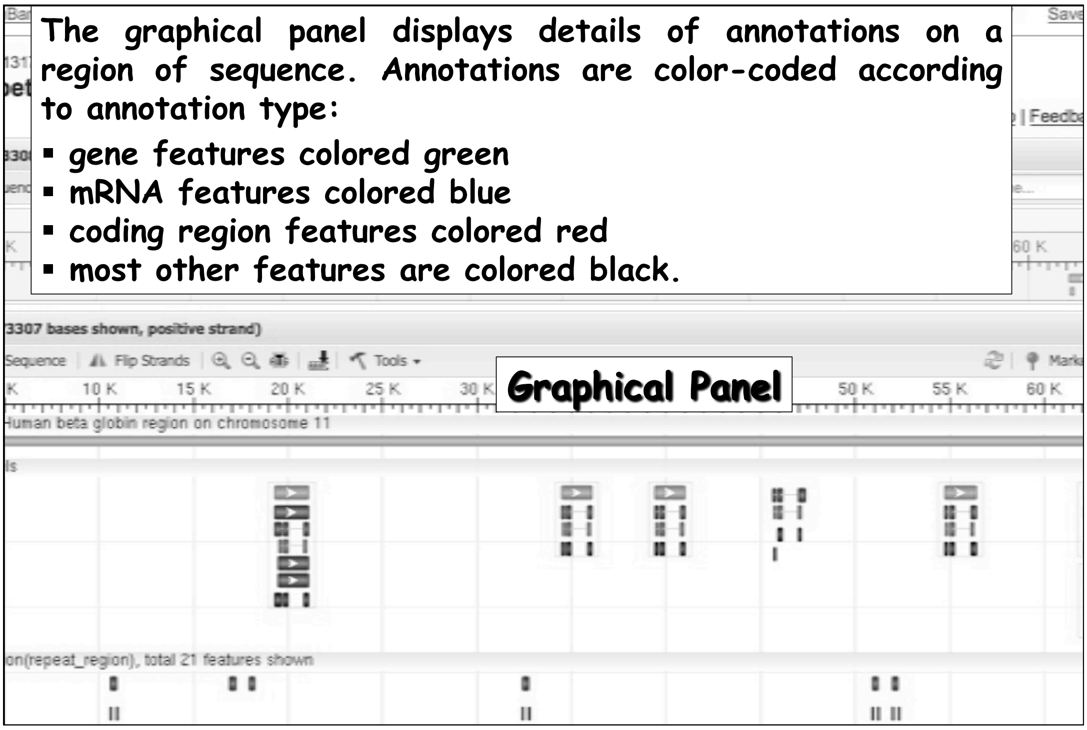
link rete



GenBank: U01317.1

Human beta globin region on chromosome 11

Overview Panel



The graphical panel displays details of annotations on a region of sequence. Annotations are color-coded according to annotation type:

- gene features colored green
- mRNA features colored blue
- coding region features colored red
- most other features are colored black.

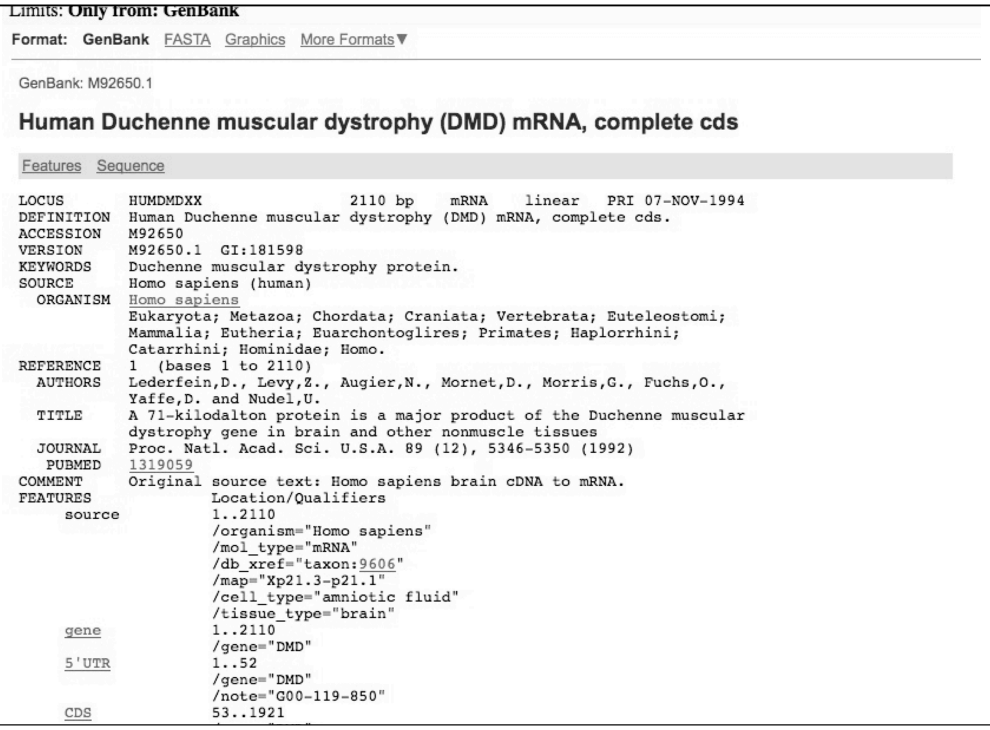
Graphical Panel

```
      12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
44890 ACCCCATCTCTACTAAAAATACAAAAAATTAGCCAGAGACGGTGGCGGGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCGTGAACCTGGGAGGCA
45000 GAGCTGCAGTGAGCAGAGATCGCGCCACTGCACACTCCAGCCTGGGTGACAAAGCGAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAGAAAGAAAGAAAGAAAAAA
45110 AAAGTGATGAAATTGTGTATTCAATGTAGTCTCAAGAGAATTGAAAACCAAGAAAGGCTGTGGCTTCTTCCACATAAAGCCTGGATGAATAACAGGATAACACGTTGTTA
45220 CATTGTCACAACTCCTGATCCAGGAATTGATGGCTAAGATATTCGTAATTCTTATCCTTTTCAGTTGTAACTTATTCCTATTTGTCAGCATTCAGGTTATTAGCGGCTGC
45330 TGGCGAAGTCCTTGAGAAATAAACTGCACACTGGATGGTGGGGTAGTGTAGGAAAATGGAGGGGAAGGAAGTAAAGTTTCAAATTAAGCCTGAACAGCAAAGTTCCCCT
45440 GAGAAGGCCACCTGGATTCTATCAGAAACTCGAATGTCCATCTTGCAAAACTTCCTTGCCCAAACCCCACCCCTGGAGTCACAACCCACCCTTGACCAATAGATTCATTT
45550 CACTGAGGGAGGCAAAGGGCTGGTCAATAGATTCATTTCACTGGGAGAGGCAAAGGGCTGGGGGCCAGAGAGGAGAAGTAAAAAGCCACACATGAAGCAGCAATGCAGGC
45660 ATGCTTCTGGCTCATCTGTGATCACCAGGAAACTCCCAGATCTGACACTGTAGTGCATTTCACTGCTGACAAGAAGGCTGCTGCCACCAGCCTGTGAAGCAAGGTTAAGG
                                                        -  V  H  F  T  A  D  K  K  A  A  A  T  S  L  *  S  K  V  K
45770 TGAGAAGGCTGGAGGTGAGATTCTGGGCAGGTAGGTACTGGAAGCCGGGACAAGGTGCAGAAAGGCAGAAAGTGTTTCTGAAAGAGGGATTAGCCCGTTGTCTTACATAG
      V  R  R  L  E  V  R  F  W  A
45880 TCTGACTTTGCACCTGCTCTGTGATTATGACTATCCCACAGTCTCCTGGTTGTCTACCCATGGACCTAGAGGTACTTTGAAAGTTTTGGATATCTGGGCTCTGACTGTGC
                                               V  S  W  L  S  T  H  G  P  R  G  T  L  K  V  L  D  I  W  A  L  T  V
45990 AATAATGGGCAACCCCAAAGTCAAGGCACATGGCAAGAAGGTGCTGATCTCCTTCGGAAAAGCTGTTATGCTCACGGATGACCTCAAAGGCACCTTTGCTACACTGAGTG
      Q  *  W  A  T  P  K  S  R  H  M  A  R  R  C  *  S  P  S  E  K  L  L  C  S  R  M  T  S  K  A  P  L  L  H  *  V
46100 ACCTGCACTGTAACAAGCTGCACGTGGACCCTGAGAACTTCCTGGTGAGTAGTAAGTACACTCACGCTTTCTTCTTTACCCTTAGATATTTGCACTATGGGTACTTTTGA
       T  C  T  V  T  S  C  T  W  T  L  R  T  S  W
46210 AAGCAGAGGTGGCTTTCTCTTGTGTTATGAGTCAGCTATGGGATATGATATTTCAGCAGTGGGATTTTGAGAGTTATGTTGCTGTAAATAACATAACTAAAATTTGGTAG
46320 AGCAAGGACTATGAATAATGGAAGGCCACTTACCATTTGATAGCTCTGAAAAACACATCTTATAAAAAATTCTGGCCAAAATCAAACTGAGTGTTTTGGATGAGGGAACA
46430 GAAGTTGAGATAGAGAAAATAACATCTTTCCTTTGGTCAGCGAAATTTTCTATAAAAATTAATAGTCACTTTTCTGCATAGTCCTGGAGGTTAGAAAAAGATCAACTGAA
46540 CAAAGTAGTGGGAAGCTGTTAAAAGAGGATTGTTTCCCTCCGAATGATGATGGTATACTTTTGTACGCATGGTACAGGATTCTTTGTTATGAGTGTTTGGGAAAATTGTA
46650 TGTATGTATGTATGTATGTGATGACTGGGGACTTATCCTATCCATTACTGTTCCTTGAAGTACTATTATCCTACTTTTTAAAAGGACGAAGTCTCTAAAAAAAAAATGAA
46760 ACAATCACAATATGTTGGGGTAGTGAGTTGGCATAGCAAGTAAGAGAAGGATAGGACACAATGGGAGGTGCAGGGCTGCCAGTCATATTGAAGCTGATATCTAGCCCATA
46870 ATGGTGAGAGTTGCTCAAACTCTGGTCAAAAAGGATGTAAGTGTTATATCTATTTACTGCAAGTCCAGCTTGAGGCCTTCTATTCACTATGTACCATTTTCTTTTTTATC
46980 TTCACTCCCTCCCCAGCTCTTAGGCAACGTGATATTGATTGTTTTGGCAACCCACTTCAGCGAGGATTTTACCCTACAGATACAGGCTTCTTGGCAGTAACTAACAAATG
                       L  L  G  N  V  I  L  I  V  L  A  T  H  F  S  E  D  F  T  L  Q  I  Q  A  S  W  Q  *  L  T  N
47090 CTGTGGTTAATGCTGTAGCCCACAAGACCACTGAGTTCCCTGTCCACTATGTTTGTACCTATGTCCCAAAATCTCATCTCCTTTAGATGGGGGAGGTTGGGGAGAAGAGC
      A  V  V  N  A  V  A  H  K  T  T
47200 AGTATCCTGCCTGCTGATTCAGTTCCTGCATGATAAAAATAGAATAAAGAAATATGCTCTCTAAGAAATATCATTGTACTCTTTTTCTGTCTTTATATTTTACCCTGATT
47310 CAGCCAAAAGGACGCACTATTTCTGATGGAAATGAGAATGTTGGAGAATGGGAGTTTAAGGACAGAGAAGATACTTTCTTGCAATCCTGCAAGAAAAGAGAGAACTCGTG
47420 GGTGGATTTAGTGGGGTAGTTACTCCTAGGAAGGGGAAATCGTCTCTAGAATAAGACAATGTTTTTACAGAAAGGGAGGTCAATGGAGGTACTCTTTGGAGGTGTAAGAG
47530 GATTGTTGGTAGTGTGTAGAGGTATGTTAGGACTCAAATTAGAAGTTCTGTATAGGCTATTATTTGTATGAAACTCAGGATATAGCTCATTTGGTGACTGCAGTTCACTT
47640 CTACTTATTTTAAACAACATATTTTTTATGATTTATAATGAAGTGGGGATGGGGCTTCCTAGAGACCAATCAAGGGCCAAACCTTGAACTTTCTCTTAACGTCTTCAATG
47750 GTATTAATAGAGAATTATCTCTAAGGCATGTGAACTGGCTGTCTTGGTTTTCATCTGTACTTCATCTGCTACCTCTGTGACCTGAAACATATTTATAATTCCATTAAGCT
47860 GTGCATATGATAGATTTATCATATGTATTTTCCTTAAAGGATTTTTGTAAGAACTAATTGAATTGATACCTGTAAAGTCTTTATCACACTACCCAATAAATAATAAATCT
47970 CTTTGTTCAGCTCTCTGTTTCTATAAATATGTACAAGTTTTATTGTTTTTAGTGGTAGTGATTTTATTCTCTTTCTATATATATACACACACATGTGTGCATTCATAAAT
```

```
     gene            34478..36069
                     /gene="HBG2"
     mRNA            join(34478..34622,34745..34967,35854..36069)
                     /gene="HBG2"
                     /product="G-gamma globin"
     exon            34478..34622
                     /gene="HBG2"
                     /note="G-gamma globin"
                     /number=1
     CDS             join(34531..34622,34745..34967,35854..35982)
                     /gene="HBG2"
                     /codon_start=1
                     /product="G-gamma globin"
                     /protein_id="AAA16331.1"
                     /db_xref="GI:455994"
                     /translation="MGHFTEEDKATITSLWGKVNVEDAGGETLGRLLVVYPWTQRFFD
                     SFGNLSSASAIMGNPKVKAHGKKVLTSLGDAIKHLDDLKGTFAQLSELHCDKLHVDPE
                     NFKLLGNVLVTVLAIHFGKEFTPEVQASWQKMVTGVASALSSRYH"
     exon            34745..34967
                     /gene="HBG2"
                     /number=2
     exon            35854..36069
                     /gene="HBG2"
                     /note="G-gamma globin"
                     /number=3
     conflict        37486
                     /citation=[29]
                     /replace="t"
     variation       37657..40179
                     /note="g-25 kb-c in wt; gc in silent deletion [91]"
     variation       37975
                     /note="t in all Benin-type sickle cell (beta-S)
                     chromosomes and some African beta-A chromosomes; t
                     abolishes a TaqI site"
                     /citation=[29]
                     /replace="t"
     variation       39297
                     /note="g in wt (b allele, see separate sequence); a in
                     HPFH [81],[82]"
```

http://www.ncbi.nlm.nih.gov/nuccore/

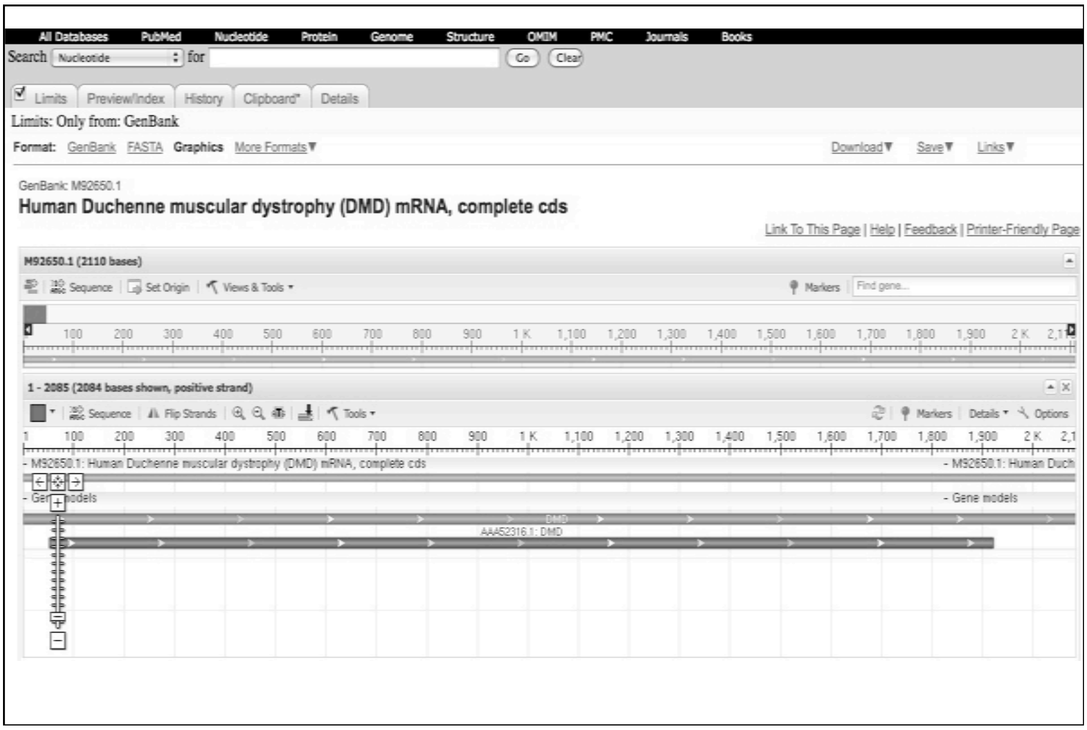
link rete



GenBank: M92650.1

Human Duchenne muscular dystrophy (DMD) mRNA, complete cds

```
LOCUS       HUMDMDXX                2110 bp    mRNA    linear   PRI 07-NOV-1994
DEFINITION  Human Duchenne muscular dystrophy (DMD) mRNA, complete cds.
ACCESSION   M92650
VERSION     M92650.1  GI:181598
KEYWORDS    Duchenne muscular dystrophy protein.
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 2110)
  AUTHORS   Lederfein,D., Levy,Z., Augier,N., Mornet,D., Morris,G., Fuchs,O.,
            Yaffe,D. and Nudel,U.
  TITLE     A 71-kilodalton protein is a major product of the Duchenne muscular
            dystrophy gene in brain and other nonmuscle tissues
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (12), 5346-5350 (1992)
   PUBMED   1319059
COMMENT     Original source text: Homo sapiens brain cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1..2110
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="Xp21.3-p21.1"
                     /cell_type="amniotic fluid"
                     /tissue_type="brain"
     gene            1..2110
                     /gene="DMD"
     5'UTR           1..52
                     /gene="DMD"
                     /note="G00-119-850"
     CDS             53..1921
```



Limits: Only from: GenBank

Format: GenBank FASTA **Graphics** More Formats

GenBank: M92650.1

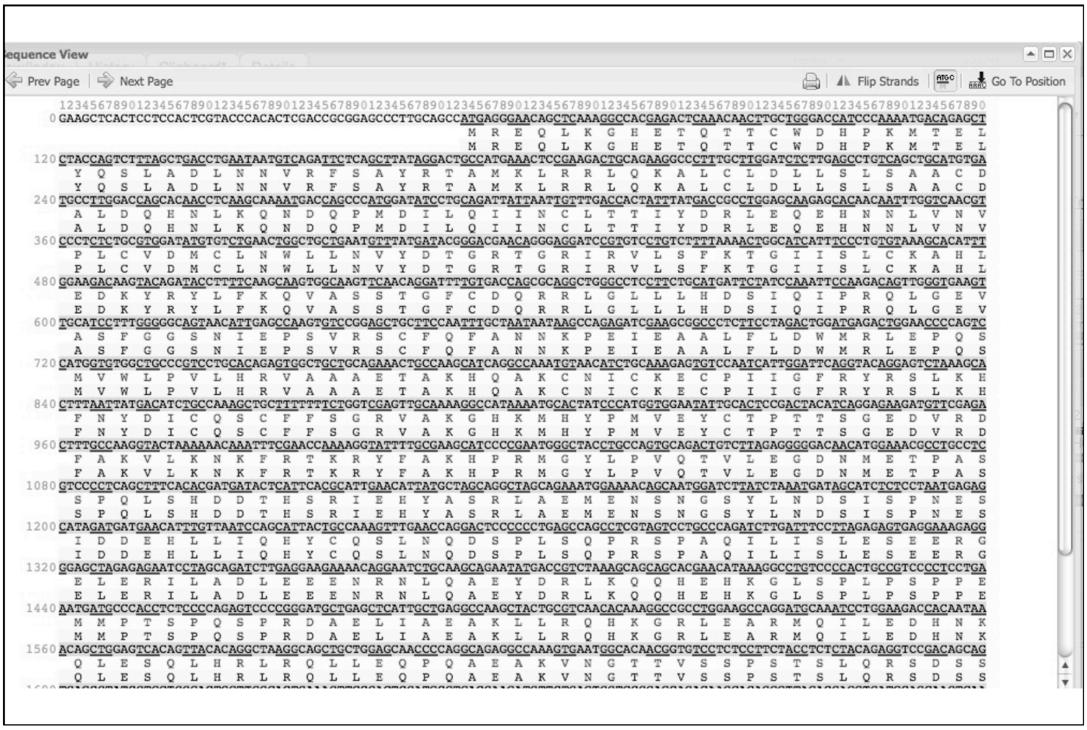
Human Duchenne muscular dystrophy (DMD) mRNA, complete cds

M92650.1 (2110 bases)

1 - 2085 (2084 bases shown, positive strand)

- M92650.1: Human Duchenne muscular dystrophy (DMD) mRNA, complete cds

- Gene models



```
     123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
   0 GAAGCTCACTCCTCCACTCGTACCCACACTCGACCGCGGAGCCCTTGCAGCCATGAGGGAACAGCTCAAAGGCCACGAGACTCAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCT
                                                           M  R  E  Q  L  K  G  H  E  T  Q  T  T  C  W  D  H  P  K  M  T  E  L
                                                           M  R  E  Q  L  K  G  H  E  T  Q  T  T  C  W  D  H  P  K  M  T  E  L
 120 CTACCAGTCTTTAGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGA
       Y  Q  S  L  A  D  L  N  N  V  R  F  S  A  Y  R  T  A  M  K  L  R  R  L  Q  K  A  L  C  L  D  L  L  S  L  S  A  A  C  D
       Y  Q  S  L  A  D  L  N  N  V  R  F  S  A  Y  R  T  A  M  K  L  R  R  L  Q  K  A  L  C  L  D  L  L  S  L  S  A  A  C  D
 240 TGCCTTGGACCAGCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGT
       A  L  D  Q  H  N  L  K  Q  N  D  Q  P  M  D  I  L  Q  I  I  N  C  L  T  T  I  Y  D  R  L  E  Q  E  H  N  N  L  V  N  V
       A  L  D  Q  H  N  L  K  Q  N  D  Q  P  M  D  I  L  Q  I  I  N  C  L  T  T  I  Y  D  R  L  E  Q  E  H  N  N  L  V  N  V
 360 CCCTCTCTGCGTGGATATGTGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTT
       P  L  C  V  D  M  C  L  N  W  L  L  N  V  Y  D  T  G  R  T  G  R  I  R  V  L  S  F  K  T  G  I  I  S  L  C  K  A  H  L
       P  L  C  V  D  M  C  L  N  W  L  L  N  V  Y  D  T  G  R  T  G  R  I  R  V  L  S  F  K  T  G  I  I  S  L  C  K  A  H  L
 480 GGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGGCCTCCTTCTGCATGATTCTATCCAAATTCCAAGACAGTTGGGTGAAGT
       E  D  K  Y  R  Y  L  F  K  Q  V  A  S  S  T  G  F  C  D  Q  R  R  L  G  L  L  L  H  D  S  I  Q  I  P  R  Q  L  G  E  V
       E  D  K  Y  R  Y  L  F  K  Q  V  A  S  S  T  G  F  C  D  Q  R  R  L  G  L  L  L  H  D  S  I  Q  I  P  R  Q  L  G  E  V
 600 TGCATCCTTTGGGGGCAGTAACATTGAGCCAAGTGTCCGGAGCTGCTTCCAATTTGCTAATAATAAGCCAGAGATCGAAGCGGCCCTCTTCCTAGACTGGATGAGACTGGAACCCCAGTC
       A  S  F  G  G  S  N  I  E  P  S  V  R  S  C  F  Q  F  A  N  N  K  P  E  I  E  A  A  L  F  L  D  W  M  R  L  E  P  Q  S
       A  S  F  G  G  S  N  I  E  P  S  V  R  S  C  F  Q  F  A  N  N  K  P  E  I  E  A  A  L  F  L  D  W  M  R  L  E  P  Q  S
 720 CATGGTGTGGCTGCCCGTCCTGCACAGAGTGGCTGCTGCAGAAACTGCCAAGCATCAGGCCAAATGTAACATCTGCAAAGAGTGTCCAATCATTGGATTCAGGTACAGGAGTCTAAAGCA
       M  V  W  L  P  V  L  H  R  V  A  A  A  E  T  A  K  H  Q  A  K  C  N  I  C  K  E  C  P  I  I  G  F  R  Y  R  S  L  K  H
       M  V  W  L  P  V  L  H  R  V  A  A  A  E  T  A  K  H  Q  A  K  C  N  I  C  K  E  C  P  I  I  G  F  R  Y  R  S  L  K  H
 840 CTTTAATTATGACATCTGCCAAAGCTGCTTTTTTTCTGGTCGAGTTGCAAAAGGCCATAAAATGCACTATCCCATGGTGGAATATTGCACTCCGACTACATCAGGAGAAGATGTTCGAGA
       F  N  Y  D  I  C  Q  S  C  F  F  S  G  R  V  A  K  G  H  K  M  H  Y  P  M  V  E  Y  C  T  P  T  T  S  G  E  D  V  R  D
       F  N  Y  D  I  C  Q  S  C  F  F  S  G  R  V  A  K  G  H  K  M  H  Y  P  M  V  E  Y  C  T  P  T  T  S  G  E  D  V  R  D
 960 CTTTGCCAAGGTACTAAAAAACAAATTTCGAACCAAAAGGTATTTTGCGAAGCATCCCCGAATGGGCTACCTGCCAGTGCAGACTGTCTTAGAGGGGGACAACATGGAAACGCCTGCCTC
       F  A  K  V  L  K  N  K  F  R  T  K  R  Y  F  A  K  H  P  R  M  G  Y  L  P  V  Q  T  V  L  E  G  D  N  M  E  T  P  A  S
       F  A  K  V  L  K  N  K  F  R  T  K  R  Y  F  A  K  H  P  R  M  G  Y  L  P  V  Q  T  V  L  E  G  D  N  M  E  T  P  A  S
1080 GTCCCCTCAGCTTTCACACGATGATACTCATTCACGCATTGAACATTATGCTAGCAGGCTAGCAGAAATGGAAAACAGCAATGGATCTTATCTAAATGATAGCATCTCTCCTAATGAGAG
       S  P  Q  L  S  H  D  D  T  H  S  R  I  E  H  Y  A  S  R  L  A  E  M  E  N  S  N  G  S  Y  L  N  D  S  I  S  P  N  E  S
       S  P  Q  L  S  H  D  D  T  H  S  R  I  E  H  Y  A  S  R  L  A  E  M  E  N  S  N  G  S  Y  L  N  D  S  I  S  P  N  E  S
1200 CATAGATGATGAACATTTGTTAATCCAGCATTACTGCCAAAGTTTGAACCAGGACTCCCCCCTGAGCCAGCCTCGTAGTCCTGCCCAGATCTTGATTTCCTTAGAGAGTGAGGAAAGAGG
       I  D  D  E  H  L  L  I  Q  H  Y  C  Q  S  L  N  Q  D  S  P  L  S  Q  P  R  S  P  A  Q  I  L  I  S  L  E  S  E  E  R  G
       I  D  D  E  H  L  L  I  Q  H  Y  C  Q  S  L  N  Q  D  S  P  L  S  Q  P  R  S  P  A  Q  I  L  I  S  L  E  S  E  E  R  G
1320 GGAGCTAGAGAGAATCCTAGCAGATCTTGAGGAAGAAAACAGGAATCTGCAAGCAGAATATGACCGTCTAAAGCAGCAGCACGAACATAAAGGCCTGTCCCCACTGCCGTCCCCTCCTGA
       E  L  E  R  I  L  A  D  L  E  E  E  N  R  N  L  Q  A  E  Y  D  R  L  K  Q  Q  H  E  H  K  G  L  S  P  L  P  S  P  P  E
       E  L  E  R  I  L  A  D  L  E  E  E  N  R  N  L  Q  A  E  Y  D  R  L  K  Q  Q  H  E  H  K  G  L  S  P  L  P  S  P  P  E
1440 AATGATGCCCACCTCTCCCCAGAGTCCCCGGGATGCTGAGCTCATTGCTGAGGCCAAGCTACTGCGTCAACACAAAGGCCGCCTGGAAGCCAGGATGCAAATCCTGGAAGACCACAATAA
       M  M  P  T  S  P  Q  S  P  R  D  A  E  L  I  A  E  A  K  L  L  R  Q  H  K  G  R  L  E  A  R  M  Q  I  L  E  D  H  N  K
       M  M  P  T  S  P  Q  S  P  R  D  A  E  L  I  A  E  A  K  L  L  R  Q  H  K  G  R  L  E  A  R  M  Q  I  L  E  D  H  N  K
1560 ACAGCTGGAGTCACAGTTACACAGGCTAAGGCAGCTGCTGGAGCAACCCCAGGCAGAGGCCAAAGTGAATGGCACAACGGTGTCCTCTCCTTCTACCTCTCTACAGAGGTCCGACAGCAG
       Q  L  E  S  Q  L  H  R  L  R  Q  L  L  E  Q  P  Q  A  E  A  K  V  N  G  T  T  V  S  S  P  S  T  S  L  Q  R  S  D  S  S
       Q  L  E  S  Q  L  H  R  L  R  Q  L  L  E  Q  P  Q  A  E  A  K  V  N  G  T  T  V  S  S  P  S  T  S  L  Q  R  S  D  S  S
```