

# Πληροφοριακά συστήματα Γονιδιώματος

## GenBank

- Η GenBank είναι μία βάση δεδομένων με πληροφορίες DNA. Την βάση την διαχειρίζεται το National Center for Biotechnology Information (NCBI). Η GenBank χωρίζεται σε 17 διακριτές κατηγορίες (divisions) ανάλογα με το είδος της αλληλουχίας DNA που περιέχουν:

| <b>Division</b> | <b>sequence subset</b> | <b>Division</b> | <b>sequence subset</b>       |
|-----------------|------------------------|-----------------|------------------------------|
| • PRI           | primate                | PHG             | Bacteriophage                |
| • ROD           | rodent                 | SYN             | synthetic                    |
| • MAM           | other mammalian        | UNA             | unannotated                  |
| • VRT           | other vertebrate       | EST             | expressed sequence tags      |
| • INV           | invertebrate           | PAT             | patent                       |
| • PLN           | plant, fungal, algal   | STS             | sequence tagged sites        |
| • BCT           | bacterial              | GSS             | genome survey sequences      |
| • RNA           | structural RNA         | HTG             | high throughput genomic seq. |
| • VRL           | viral                  |                 |                              |

- Πληροφορίες μπορεί κανείς να ανακτήσει από τη GenBank, χρησιμοποιώντας το ολοκληρωμένο σύστημα ανάκτησης πληροφοριών Entrez.

## **Δομή των καταχωρήσεων της GenBank**

- Μια καταχώρηση στη GenBank περιέχει το αρχείο αλληλουχίας, το οποίο περιέχει εκτός από την αλληλουχία καθ'αυτή και διάφορες περιγραφικές πληροφορίες που σχετίζονται με αυτήν.
- Κάθε καταχώρηση (entry) επίσης περιλαμβάνει και διάφορες λέξεις-κλειδιά, σχετιζόμενα υπό-κλειδιά και ένα προαιρετικό αρχείο αξιοσημείων χαρακτηριστικών(feature table).
- Μια GenBank γραμμογράφηση μιας καταχώρηση για την ανθρώπινη cyclooxygenase (<http://en.wikipedia.org/wiki/Cyclooxygenase>)είναι η ακόλουθη:

- LOCUS HUMCYCLOX 3387 bp mRNA linear PRI 31-DEC-1994
- DEFINITION Homo sapiens cyclooxygenase-2 (Cox-2) mRNA, complete cds.
- ACCESSION M90100
- VERSION M90100.1 GI:181253
- KEYWORDS cyclooxygenase-2; prostaglandin synthase.
- SOURCE Homo sapiens (human)
- ORGANISM [Homo sapiens](#)
  - Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  - Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
- REFERENCE 1 (bases 1 to 3387)
- AUTHORS Hla,T. and Neilson,K.
- TITLE Human cyclooxygenase-2 cDNA
- JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (16), 7384-7388 (1992)
- MEDLINE [92366465](#)
- PUBMED [1380156](#)
- COMMENT Original source text: Homo sapiens umbilical vein cDNA to mRNA.

- FEATURES            Location/Qualifiers
- source            1..3387
- /organism="Homo sapiens"
- /mol\_type="mRNA"
- /db\_xref="taxon:9606"
- /cell\_type="endothelial"
- /tissue\_type="umbilical vein"
- [gene](#)                1..3387
- /gene="Cox-2"
- [5'UTR](#)            1..97
- /gene="Cox-2"
- [CDS](#)                98..1912
- /gene="Cox-2"
- /EC\_number="[1.14.99.1](#)"
- /codon\_start=1
- /product="cyclooxygenase-2"
- /protein\_id="[AAA58433.1](#)"
- /db\_xref="GI:181254"
- /translation="MLARALLLCAVLALSHTANPCCSHPCQNRGVCMSVGFQDKKDC
- TRTGFYGENCSTPEFLTRIKLFLKPTPNTVHYILTHFKGFVNVVNNIPFLRNAIMSYV
- ...
- VEVGAPFSLKGLMGNVICSPAYWKPSTFGGEVGFQIINTASIQSLICNNVKGCPFTSF
- SVPDPELIKVTINASSRSGLDDINPTVLLKERSTEL"
- [sig\\_peptide](#)    98..148
- /gene="Cox-2"
- [mat\\_peptide](#)    149..1909
- /gene="Cox-2"
- /product="cyclooxygenase-2"
- /EC\_number="[1.14.99.1](#)"
- [3'UTR](#)            1913..3387
- /gene="Cox-2"
- [polyA\\_signal](#)    3369..3374
- /gene="Cox-2"

- ORIGIN
- 1 gtccaggaac tcctcagcag cgcctccttc agtccacag ccagacgccc tcagacagca
- 61 aagcctaccc cgcgcccgc cctgcccgc cgctgcatg ctcgcccgc cctgctgt
- 121 gtgcgcggtc ctggcgctca gccatacagc aaatcctgc tgtcccacc catgtcaaaa
- 181 ccgaggtgta tgtatgagtg tgggattga ccagtataag tgcgattgta cccggacagg
- 241 attctatgga gaaaactgct caacaccgga attttgaca agaataaaat tatttctgaa
- 301 acccactcca aacacagtgc actacatact taccacttc aagggatttt ggaacgttgt
- 361 gaataacatt ccctccttc gaaatgcaat tatgagttat gtgtgacat ccagatcaca
- ...
- 3301 tacctgaact ttgcaagtt ttcaggtaaa cctcagctca ggactgctat ttagctcctc
- 3361 ttaagaagat taaaaaaaaa aaaaaaag
- //

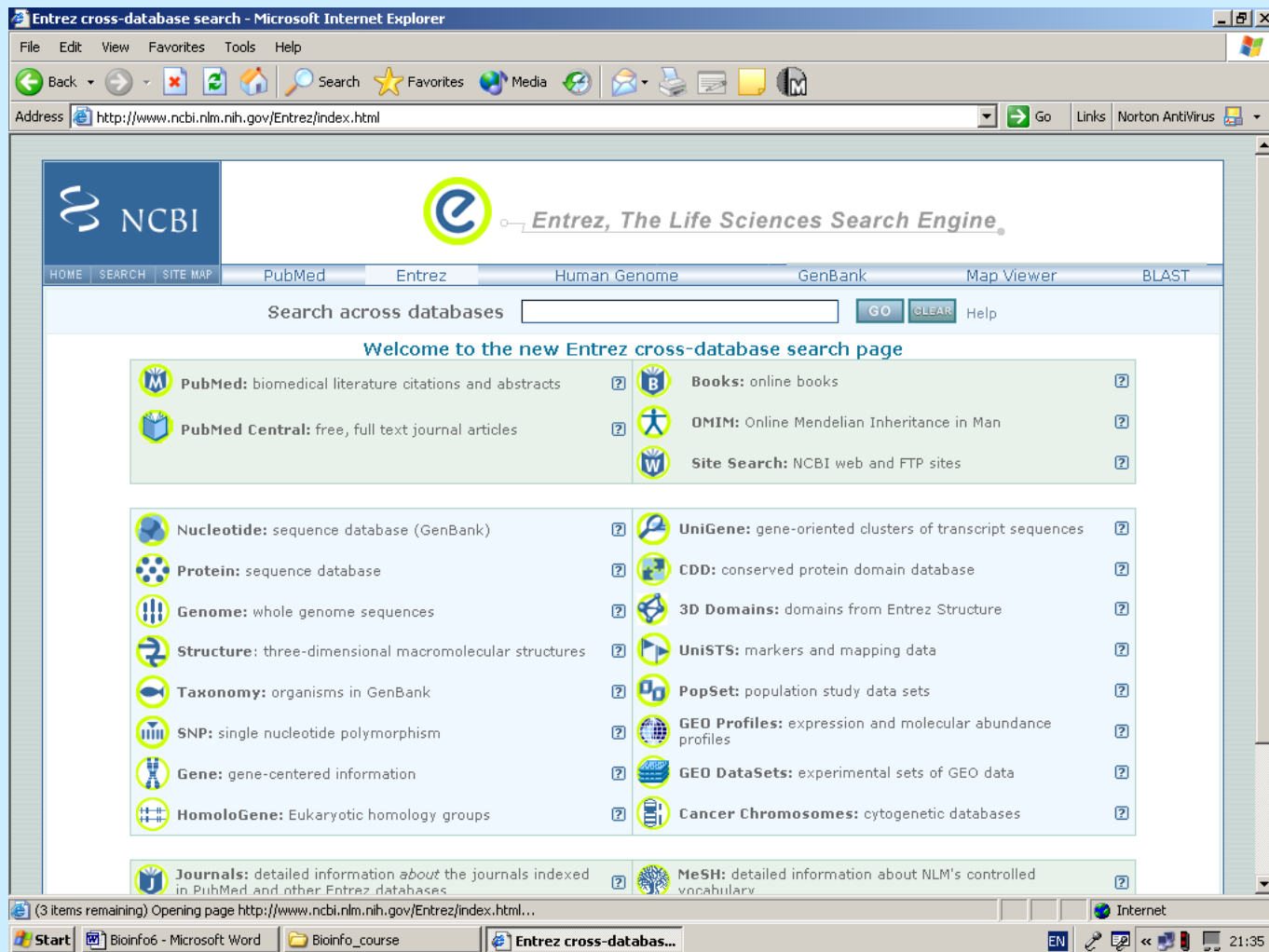
- Η Γραμμή LOCUS είναι το χαρακτηριστικό που υποδουλώνει αλληλουχία λειτουργικότητας (HUMCYCLOX υποδουλώνει ανθρώπινη cyclooxygenase), τον αριθμό βάσεων, την προέλευση της πληροφορίας αλληλουχίας (mRNA), το τμήμα της ΒΠ (PRI) και την ημερομηνία υποβολής των δεδομένων στη ΒΠ.
- Η γραμμή DEFINITION περιέχει την περιγραφή της αλληλουχίας.
- Η γραμμή ACCESSION είναι ένας μοναδικός κωδικός ο οποίος δίνεται σε κάθε μια καταχώρηση.
- Η γραμμή KEYWORDS περιέχει φράσεις που περιγράφουν τα γονίδια και άλλες σχετικές πληροφορίες.

- Η γραμμή SOURCE περιέχει πληροφορίες για τον ιστό από τον οποίο τα δεδομένα εξήχθησαν.
- Η γραμμή ORGANISM περιγράφει την βιολογική ταξινόμηση του οργανισμού προέλευσης.
- Η γραμμή REFERENCE παρέχει παραπομπές στη βιβλιογραφία για την συγκεκριμένη αλληλουχία.
- Η γραμμή FEATURES παρέχει ακριβείς πληροφορίες για τα χαρακτηριστικά (Feature Table). Συντεταγμένες παρέχονται για το 5' μη-μεταφρασμένο μέρος (1-97), για την αλληλουχία κωδικοποίησης (98-1912), για το 3' μη-μεταγραφόμενο μέρος (1913-3387), για τη polyadenylation αλληλουχία (3369-3374) κ.τ.λ. Επίσης παρέχονται πληροφορίες για τη μετάφραση της πρωτεΐνης και τις θέσεις διαφόρων πεπτιδίων.
- Η γραμμή BASE COUNT δίνει πληροφορίες για την συχνότητα εμφάνισης των διαφορετικών τύπων βάσεων στην αλληλουχία (i.e. 1010 A, 712 C, 633 G, 1032 T).
- Η γραμμή ORIGIN σημειώνει την πρώτη βάση της αλληλουχίας στο γονιδίωμα.

- ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ

- **GenBank**

- Για να προσπελάσετε τη ΒΠ GenBank, στον Internet Explorer πληκτρολογήστε την ακόλουθη διεύθυνση : [www.ncbi.nlm.nih.gov/Entrez/index.html](http://www.ncbi.nlm.nih.gov/Entrez/index.html). Στη συνέχεια επιλέξτε “Nucleotide: sequence database (GenBank)”.





- Για να ανακτήσετε πληροφορίες για το HUMCYCLOX, πληκτρολογήστε τον Accession Code στο πλαίσιο παρακάτω και επιλέξτε “GO”.

The screenshot shows the Entrez Nucleotide search interface. The browser window title is "Entrez Nucleotide - Microsoft Internet Explorer". The address bar contains the URL: `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide`. The search bar is set to "Nucleotide" and contains the query "HUMCYCLOX". The search results area displays a yellow highlight with the text: "The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases grows at an exponential rate. As of April 2004, there are over 38,989,342,565 bases." Below this, there is a section titled "Human Genome" with the text: "Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#)." Further down, there is a section titled "Building the human genome" with the text: "The Human Genome Reference DNA Sequence was completed in April 2003. The current version is listed as a build number on the [Genome View](#) page and includes an accompanying set of [statistics](#) and [release notes](#)." At the bottom, there is a section titled "Homo sapiens genome view" with the text: "build 34 version 1 statistics" and a diagram of the human genome.

Στη συνέχεια επιλέξτε “M90100”, τον accession code για Homo sapiens cyclooxygenase-2 (Cox-2) mRNA.

The screenshot shows a Microsoft Internet Explorer browser window displaying the NCBI Entrez Nucleotide search results. The address bar contains the URL: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=nucleotide>. The search query is "HUMCYCLOX" and the results are displayed in a table format. The first result is "M90100" with the description "Homo sapiens cyclooxygenase-2 (Cox-2) mRNA, complete cds" and the accession code "gi|181253|gb|M90100.1|HUMCYCLOX[181253]". The browser window also shows the NCBI logo and various navigation options like "Limits", "Preview/Index", "History", "Clipboard", and "Details".

| Accession              | Description   |
|------------------------|---|
| <a href="#">M90100</a> | Homo sapiens cyclooxygenase-2 (Cox-2) mRNA, complete cds<br>gi 181253 gb M90100.1 HUMCYCLOX[181253] |

- Το σύστημα θα παρουσιάσει την καταχώρηση. Μπορείτε να κυλήσετε την οθόνη προς τα κάτω για να δείτε διάφορες πληροφορίες καθώς και την αλληλουχία.

NCBI Sequence Viewer - Microsoft Internet Explorer

Address: <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=181253>

NCBI

Search  for

Display  Show:

1: [M90100](#). Homo sapiens cycl...[gi:181253] Links

LOCUS HUMCYCLOX 3387 bp mRNA linear PRI 31-DEC-1994

DEFINITION Homo sapiens cyclooxygenase-2 (Cox-2) mRNA, complete cds.

ACCESSION M90100

VERSION M90100.1 GI:181253

KEYWORDS cyclooxygenase-2; prostaglandin synthase.

SOURCE Homo sapiens (human)

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

REFERENCE 1 (bases 1 to 3387)

AUTHORS Hla, T. and Neilson, K.

TITLE Human cyclooxygenase-2 cDNA

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (16), 7384-7388 (1992)

MEDLINE [92366465](#)

PUBMED [1380156](#)

COMMENT Original source text: Homo sapiens umbilical vein cDNA to mRNA.

FEATURES

Location/Qualifiers

source

1..3387

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/cell\_type="endothelial"

/tissue\_type="umbilical vein"

[gene](#)

1..3387

/gene="Cox-2"

[5' UTR](#)

1..97

/gene="Cox-2"

Done Internet

Start Bioinfo6 - Microsoft Word Bioinfo\_course NCBI Sequence View... EN 21:49

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=181253> Go Links Norton AntiVirus

```

/tissue_type="umbilical vein"
gene 1..3387
      /gene="Cox-2"
5' UTR 1..97
      /gene="Cox-2"
CDS 98..1912
     /gene="Cox-2"
     /EC_number="1.14.99.1"
     /codon_start=1
     /product="cyclooxygenase-2"
     /protein_id="AAAS8433.1"
     /db_xref="GI:181254"
     /translation="MLARALLLCAVLALSHTANPCSSHPCQNRGVCMSVGFQYKDC
TRTGFYGENCSTPEFLTRIKLFLKPTPNTVHYILTHFKGFVNVVNNIPFLRNAIMSYV
LTSRSHLIDSPPTYNADYGYSWEAFSNLSYYTRALPPVDDCPTPLGVKGGKQLPDS
NEIVGKLLLRKFIPDPQGSNMMFAFFAQHFTHQFFKTDHKRGPATNGLGHGVDLNI
IYGETLARQKRLRFKDGKMKYQIIDGEMYPPTVKDTQAEMIYPPQVPEHLRFVVGQE
VFGLPGLMMYATIWLREHNRVCDVLKQEHPEWGDQLFQTSRLILIGETIKIVIEDY
VQHLSGYHFKLKFDPPELLFNKQFYQNRIAAEFNTLYHWHPLLPDPTFIHQKYNYYQ
FIYNNSILLEHGITQFVESFTRQIAGRVAAGGRNVPPAVQKVSQASIDQSRQMKYQSFN
EYRKRFLMKPYESFEELTGEKEMSALEALYGDIDAVELYPALLVEKPRPDAIFGETM
VEVGAPFSLKGLMGNVICSPAYWKPSTFGGEVGFQIINTASIQSLICNVKGCPTSTF
SVPDPELIKTVTINASSRSRGLDDINPTVLLKERSTEL"
sig_peptide 98..148
             /gene="Cox-2"
mat_peptide 149..1909
             /gene="Cox-2"
             /product="cyclooxygenase-2"
             /EC_number="1.14.99.1"
3' UTR 1913..3387
       /gene="Cox-2"
polyA_signal 3369..3374
             /gene="Cox-2"

ORIGIN
1  gtccaggaac  tcctcagcag  cgccctcctc  agctccacag  ccagacgccc  tcagacagca
61  aagcctaccc  ccgcgcgcgc  cctgcccgcg  cgtgcgatg  ctgcgccgcg  cctgctgtct
121  gtgcgcggtc  ctggcgctca  gccatacagc  aaatccttgc  tgtcccacc  catgtcaaaa
181  ccgaggtgta  tgtatgagtg  tgggatttga  ccagtataag  tgcgatttga  cccggacagg

```

Done Internet

Start Bioinfo6 - Microsoft Word Bioinfo\_course NCBI Sequence View... EN 21:49

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=181253> Go Links Norton AntiVirus

ORIGIN

```
1 gtcacaggaac tcctcagcag cgcctccttc agctccacag ccagacgccc tcagacagca
61 aagcctaccc ccgcgcgcgg ccttgcccgc cgctgcgatg ctgcgcccgcg cctctgtgct
121 gtgcgcggtc ctggcgctca gccatcacgc aaatccttgc tgttcccacc catgtcaaaa
181 ccgaggtgta tgtatgagtg tgggatttga ccagtataag tgcgattgta cccggacagg
241 attctatgga gaaaactgct caacaccgga atttttgaca agaataaaat tatttctgaa
301 acccactcca aacacagtgc actacatact taccacttc aagggatttt ggaacgttgt
361 gaataacatt cccttccttc gaaatgcaat tatgagttat gtgttgacat ccagatcaca
421 tttgattgac agtcccacaa cttacaatgc tgactatggc tacaaaagct gggaaagcctt
481 ctctaaccctc tcctattata cttagagccct tcctcctgtg cctgatgatt gcccgactcc
541 cttgggtgtc aaaggtaaaa agcagcttcc tgattcaaat gagattgtgg gaaaaattgtc
601 tctaagaaga aagtccatcc ctgatcccca gggctcaaac atgatgttg catctcttgc
661 ccagcacttc acgcatcagt ttttcaagc agatcataag cgagggccag ctttaccaca
721 cgggctgggc catgggggtgg acttaaatca tatttaccgt gaaactctgg ctagacagcg
781 taaactgcgc cttttcaagg atggaaaaat gaaatatcag ataattgatg gagagatgta
841 tcctcccaca gtcaaaagata ctcaggcaga gatgatctac cctcctcaag tccttgagca
901 tctacggttt gctggtgggc aggaggtctt tggctggtg cctggtctga tgatgtatgc
961 cacaatctgg ctgagggaac acaacagagt atgctgatgt cttaaacagg agcatcctga
1021 atggggtgat gagcagttgt tccagacaag caggctaata ctgataggag agactattaa
1081 gattgtgatt gaagattatg tgcaaacctt gagtggctat cacttcaaac tgaatttga
1141 cccagaacta cttttcaaca aacaattcca gtaccaaaat cgtattgctg ctgaatttaa
1201 caccctctat cactggcatc cccttctgcc tgacacctt caaattcatg accagaaata
1261 caactatcaa cagtttatct acaacaactc tatattgctg gaacatggaa ttaccagtt
1321 tgttgaatca ttcaccaggc aaattgctgg cagggttgcg ggtggtagga atgttcacc
1381 cgcagtacag aaagtatcac aggcctccat tgaccagagc aggcagatga aataccagtc
1441 ttttaatgag taccgcaaac gctttatgct gaagccctat gaatcatttg aagaacttac
1501 aggagaaaag gaaatgtctg cagagttgga agcactctat ggtgacatcg atgctgtgga
1561 gctgtatcct gcccttctgg tagaaaagcc tcggccagat gccatcttgg gtgaaacct
1621 ggtagaagtt ggagcccatc tctccttgaa aggacttatg ggtaatgta tatgtctcc
1681 tgccactcgg aagccaagca cttttggtgg agaagtggtt ttc aaatca tcaactctgc
1741 ctcaattcag tctctcatc gcaataacgt gaagggctgt cccttactt catcagttgt
1801 tccagatcca gagctcatta aaacagtcac catcaatgca agttcttccc gctccggact
1861 agatgatatc aatcccacag tactactaaa agaacgttgc actgaaactgt agaagtctaa
1921 tgatcatatt tatttattta tatgaacct gtctattaat ttaattattt aataatattt
1981 atattaaact ccttatgtta cttaacatct tctgtaacag aagtcagtac tcctgttggc
2041 gagaaaaggag tcataactgt gaagactttt atgtcactac tctaagatt ttgctgttgc
2101 tgtaagttt ggaaaacagt ttttattctg ttttataaac cagagagaaa tgagtttga
2161 cgtcttttta cttgaatttc aacttatatt ataaggacga aagtaagat gtttgaatac
```

Done Internet

Start Bioinfo6 - Microsoft Word Bioinfo\_course NCBI Sequence View...

EN 21:50

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=181253> Go Links Norton AntiVirus

```
1741 ctcaattcag tctctcatct gcaataacgt gaagggctgt cctttactt cattcagtg
1801 tccagatcca gagctcatta aaacagtcac catcaatgca agttcttccc gctccggact
1861 agatgatatc aatcccacag tactactaaa agaacgttcg actgaactgt agaagtctaa
1921 tgatcatatt tatttattta tatgaaccat gtctattaat ttaattatst aataatattt
1981 atattaaact ccttatgtta cttaacatct tctgtaacag aagtcagtac tctgtgtggg
2041 gagaaaaggag tcatacttgt gaagactttt atgtcactac tctaaagatt ttgctgttgc
2101 tghtaaagttt ggaaaacagt ttttattctg ttttataaac cagagagaaa tgagttttga
2161 cgtcttttta cttgaatttc aacttatatt ataaggacga aagttaagat gtttgaatac
2221 ttaaacacta tcacaagatg ccaaaatgct gaaagttttt aactgtctga tgtttccaat
2281 gcatcttcca tgatgcatta gaagtaacta atgtttgaaa ttttaaagta cttttgggta
2341 tttttctgtc atcaaacaaa acaggtatca gtgcattatt aatgaatat ttaaataga
2401 cattaccagt aatttcatgt ctacttttta aaatcagcaa tgaacaata atttgaatt
2461 tctaaattca taggtagaaa tcacctgtaa aagcttggtt gatttcttaa agttattaaa
2521 cttgtacata taccaaaaag aagctgtctt ggatttaaat ctgtaaaatc agatgaatt
2581 ttactacaat tgcttgtaa aatattttat aagtgatggt cctttttcac caagagtata
2641 aaccttttta gtgtgactgt taaaacttcc ttttaaatca aaatgccaaa tttattaagg
2701 tgggtggagcc actgcagtgt tatctcaaaa taagaatctc ctgttgagat atccagaat
2761 ctgtttatat ggctggtaac atgtaaaaac cccataaacc cgccaaaagg ggtcctacc
2821 ttgaacataa agcaataacc aaaggagaaa agcccaaat attggttcca aatttagggt
2881 ttaaacctttt tgaagcaaac ttttttttag ccttgtgcac tgcagacctg gtactcagat
2941 tttgctatga ggttaatgaa gtaccaagct gtgcttgaat aacgatatgt tttctcagat
3001 tttctgttgt acagtttaat ttagcagctc atatcacatt gcaaaagtag caatgacctc
3061 ataaaatacc tcttcaaat gcttaaatc atttcacaca ttaattttat ctcagctctg
3121 aagccaattc agtaggtgca ttggaatcaa gcctggctac ctgcatgctg ttcctttct
3181 tttcttcttt tagccatttt gctaagagac acagtcttct caaacactc gtttctccta
3241 ttttgtttta ctagttttaa gatcagagtt cactttcttt ggactctgcc tatattttct
3301 tacctgaact tttgcaagtt ttcaggtaaa cctcagctca ggactgctat ttagctctct
3361 ttaagaagat taaaaaaaa aaaaaag
```

//

[Disclaimer](#) | [Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)

Aug 4 2004 12:36:34

Start Bioinfo6 - Microsoft Word Bioinfo\_course NCBI Sequence Viewe... Internet

EM 21:50

- Για να καταχωρήσετε μια νέα ακολουθία ή να ενημερώσετε μια ήδη υπάρχουσα, επιλέξετε “Submit to GenBank” στην ιστοσελίδα παρακάτω. Στην συνέχεια ακολουθήσετε τις οδηγίες που εμφανίζονται. Το “BankIt” είναι ένα εργαλείο που μπορείτε να χρησιμοποιήσετε για να καταχωρήσετε εύκολα και απλά μια νέα αλληλουχία. Η δομή της καταχώρησης θα πρέπει να μοιάζει με αυτή της HUMCYCLOX.

Submit to GenBank - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media

Address http://www.ncbi.nlm.nih.gov/Genbank/index.html Go Links Norton AntiVirus

NCBI **Submit to GenBank**

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search Entrez for Go

NCBI

**SITEMAP**  
Guide to NCBI resources

**Accession numbers**  
For manuscript citation

**BankIt**

**Sequin**

**SequinMacroSend**  
Upload .sqn files directly

**TBL2ASN**  
Command line program

**Special submissions**  
Genomes, batch

**Submitting Sequence Data to GenBank**

The most important source of new data for GenBank® is direct submissions from scientists. GenBank depends on its contributors to help keep the database as comprehensive, current, and accurate as possible. NCBI provides timely and accurate processing and biological review of new entries and updates to existing entries, and is ready to assist authors who have new data to submit.

**Submit now!!**

[Sequin](#)  
Stand-alone sequence submission tool

[BankIt](#)  
For quick and simple submissions

[VecScreen](#)  
Vector contamination screening tool

[dbEST](#)  
[dbGSS](#)  
[dbSTS](#)  
Submit to GenBank divisions

**Receiving an Accession Number for your Manuscript**

Most journals now expect that DNA and amino acid sequences that appear in articles will be submitted to a sequence database before publication. Soon after submission, you will receive an accession number from the database which you will be able to use in your article to refer to the sequence. Please

**GenBank**

[GenBank](#)  
Overview of the database

[Search GenBank](#)

Start Bioinfo6 - Microsoft Word Bioinfo\_course Submit to GenBank - ... EN 22:00

- Επιλέξτε “BankIt”.

**BankIt: GenBank Submissions by WWW**

PubMed Entrez BLAST OMIM Taxonomy Structure

NCBI  
SITE MAP

BankIt Help  
Getting Started

Submission Info

Reference Info

Source Info

Input DNA

Additional Info

Coding Regions

RNA Features

Other Features

VecScreen

### GenBank Direct Submission Options

Use BankIt if:

- you have one or a few sequence submissions
- you prefer to use a WWW-based submission tool
- your sequence annotation is not complicated
- you do not require sequence analysis tools to submit your sequence(s)

Use [Sequin](#) if:

- you are submitting long or complex submissions
- you are submitting mutation, phylogenetic, population, environmental, or segmented sets
- you would like graphical viewing and editing options, including the alignment editor
- you would like network access to related analytical tools

Large Sequin-created files can be submitted directly through the [SequinMacroSend](#) system.

### GenBank Sequence Submission Policy

GenBank accepts DNA/RNA sequences that have been sequenced by the submitter.

Done Internet

Start Bioinfo6\_1 - Microsoft W... Bioinfo\_course BankIt: GenBank Sub... EN 22:07



- Καταχωρίστε το μέγεθος της αλληλουχίας, π.χ. 3387, και επιλέξτε “New”. Για ενημέρωση επιλέξτε “Update”.

BankIT: GenBank Submissions by WWW - Microsoft Internet Explorer

Address <http://www.ncbi.nlm.nih.gov/BankIt/>

Release Notes  
GenBank Info

- protein only sequences
- non-biologically contiguous sequences containing internal unsequenced spacers
- sequences containing a mix of genomic and mRNA sequence represented as a single sequence
- Expressed Sequence Tag (EST) submissions (should be submitted through the [dbEST](#) system)
- Genome Survey Sequence (GSS) submissions (should be submitted through the [dbGSS](#) system)

**Note:** If your submitted sequence is identical in multiple sources (eg: different geographies/specimens/isolates/strains), then you should submit each sequence from each source as a separate submission.

► **BankIt: GenBank Submissions by WWW**

- GenBank provides [annotation examples and descriptions](#) for several types of sequence submissions.
- To prepare a **New** GenBank submission, enter the size in nucleotides of your contiguous sequence here  and press
- To **Update** an existing GenBank record via a Web form and with the ability to upload a text file directly press  Click here for more detailed information about [other options for updating](#) an existing GenBank flatfile.

Revised 28 July, 2004

Done Internet

Start Bioinfo6\_1 - Microsoft W... Bioinfo\_course BankIT: GenBank Sub... EN 22:09


- Κυλήστε την οθόνη προς τα κάτω για να δείτε όλες τις πληροφορίες που χρειάζονται για να καταχωρήσετε την αλληλουχία.

BankIt: GenBank Submissions by WWW [Bottom Help](#)

**Note:** It is **STRONGLY RECOMENDED** to use [Sequin Submission and Update Tool](#) to submit multiple and complex sequences! (Examples of multiple sequences are batch sets, mutation studies, phylogenetic sets, population sets, and segmented sets)

**Note:** If this sequence is identical in multiple sources (ie: different geographies/specimens/isolates/strains), then each sequence from each source must be a **separate submission**.

There are two phases to a BankIt submission. First provide the basic submitter, citation, and sequence information requested on this page and press the [Validate and Continue](#) button at the end to review your entry in GenBank flatfile format. After that, you will be able to add coding regions, structural RNAs, and other features by following further instructions.

 **bankit658578**

**Contents :**

1. [General Submission Information](#)
2. [Reference Information](#)
3. [Source Information](#)
4. [Input DNA Sequence](#)

General Submission Information [Top](#) [Bottom](#) [Help](#)

**Multiple Submissions Information**

If you are submitting more than one sequence at this time, please number each sequence and indicate the total number of sequences to be submitted so that we can

- Αφού έχετε συμπληρώσει όλα τα απαιτούμενα πεδία στην ιστοσελίδα, πληκτρολογήστε στο πεδίο παρακάτω την αλληλουχία, και επιλέξτε “Validate and Continue”.

BankIt -- GenBank submissions by WWW - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/BankIt/nph-bankit.cgi> Go Links Norton AntiVirus

**Important:**

- Use single letter IUPAC code, raw sequence only.
- Sequence must be at least 50 bp in length
- Sequence must be biologically contiguous and not contain any internal unknown/unsequenced spacers.

Sequence length in nucleotides:

Enter DNA sequence:

**Additional Information** [Top](#) [Bottom](#) [Help](#)

- Any sequence features, such as coding regions or structural RNAs, should be added on the next page, after you "Validate and Continue" below.
- Enter any other biological information for which there is no place on the form or any pertinent instructions that will help GenBank annotators process your submission in this field.

Save this information to your local computer.

Validate the submission and correct other errors.

Start Bioinfo6\_1 - Microsoft W... Bioinfo\_course BankIt -- GenBank su... EN 22:11