

Zaslání sekvence DNA do
primární databáze

Nejdůležitější instituce zabývající se shromažďováním biomedicínských informací

- K nejdůležitějším institucím zabývajícím se, správou dat a vývojem nástrojů pro jejich analýzu a poskytováním informací patří:
 - Evropský institut pro bioinformatiku (EBI) se sídlem v Hinxtonu v UK (<http://www.ebi.ac.uk/>),
 - Národní centrum pro biotechnologické informace (NCBI) založené původně v rámci Národní lékařské knihovny (NLM) v USA (<http://www.ncbi.nlm.nih.gov/>),
 - Centrum pro informační biologii (CIB) založené jako oddělení Národního genetického institutu (NIG) v Mishimě, Japonsko (<http://www.cib.nig.ac.jp/>).

Nejdůležitější databáze sekvencí nukleových kyselin a proteinů

- V každém ze tří hlavních bioinformatických center je spravována **genomová databáze** sekvencí nukleových kyselin a odpovídajících, z nich přeložených proteinů.
 - **EMBL Nucleotide Sequence Database** (v rámci institutu EBI) – 1980
 - **GenBank** (v rámci institutu NCBI) – 1982
 - **DDBJ** (The DNA Data Bank of Japan) - 1984
- Tři samostatné báze vznikly v důsledku potřeby rychlé dostupnosti databáze sekvencí na jednotlivých kontinentech v době, kdy ještě nebyly rozvinuté vysokorychlostní komunikační sítě.

Identifikace záznamu v primárních sekvenčních databázích

- GenBank
- EMBL
- DDBJ

- **Přístupový kód (Accession Number)**
- **číslo GI (GenBank Identifier)**

```
LOCUS          AY870395                553 bp    DNA     linear   BCT 30-JAN-2005
DEFINITION    Macrococcus brunensis strain CCM 4811 60 kDa chaperonin (cpn60)
              gene, partial cds.
ACCESSION     AY870395 ←
VERSION       AY870395.1  GI:58119461 ←
```

Jak se data dostanou do databází?

- Předání dat prostřednictvím WWW
 - BankIt (GenBank)
 - <http://www.ncbi.nlm.nih.gov/BankIt/>
 - WebIn (EMBL)
 - <http://www.ebi.ac.uk/embl/Submission/webin.html>
 - Sakura (DDBJ)
 - <http://sakura.ddbj.nig.ac.jp/>
- Samostatná aplikace pro PC
 - Sequin
 - http://www.ncbi.nlm.nih.gov/Sequin/download/seq_download.html
 - pro delší sekvence (genomy)
 - fylogenetické, populační nebo mutační studie obsahující sekvenční příložen
- TPA (Third Party Annotation) anotace třetí stranou
 - záznamy, které upřesňují existující sekvence uložené do databází jinými autory
 - striktní požadavek na přímý experimentální důkaz navrhované anotace

New BankIt


BankIt - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/WebSub/?form=history&tool=

Soubor Úpravy Zobrazit Oblíbené položky Nástroje Nápověda

Oblíbené položky BankIt

Stránka Zabezpečení Nástroje

 **New BankIt** Logged in as Roman Pantucek (roman.pantucek) [Log out](#)

[Home](#) [Search](#) [Site Map](#)




Submissions

[New Submission](#)

Complete Submissions

ID	Date	Submitted Record
1391012	15 Sep 2010 10:35:52	Download File (*.zip)

[Contact](#) | [Copyright](#) | [Disclaimer](#) | [Privacy](#) | [Accessibility](#)
National Center for Biotechnology Information, US National Library of Medicine
8600 Rockville Pike, Bethesda, MD USA 20894

http://www.ncbi.nlm.nih.gov/WebSub/index.cgi?tool= Internet 100%

Old BankIt

The screenshot shows a Windows Internet Explorer browser window displaying the 'Old BankIt' website. The browser's address bar shows the URL <http://www.ncbi.nlm.nih.gov/BankIt/oldbankit.html>. The page title is 'BankIt: GenBank Submissions by WWW'. The navigation menu includes 'PubMed', 'Entrez', 'BLAST', 'OMIM', 'Taxonomy', and 'Structure'. The main content area features a yellow highlighted note about a new version of BankIt, followed by 'GenBank Direct Submission Options' and 'Sequin' options. The left sidebar contains a navigation menu with links like 'NCBI', 'SITE MAP', 'BankIt Help', 'Getting Started', 'Submission Info', 'Reference Info', 'Source Info', 'Input DNA', 'Additional Info', 'Coding Regions', 'RNA Features', 'Other Features', and 'VecScreen'. The Windows taskbar at the bottom shows the 'Internet' icon and a 75% zoom level.

BankIt: GenBank Submissions by WWW - Windows Internet Explorer

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

BankIt

Oblíbené položky BankIt: GenBank Submissi...

NCBI **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

Note: NCBI is testing a new version of the BankIt sequence submission tool, which will replace the current BankIt. This new version can be used to submit simple sets, such as single gene/CDS or rRNA/ITS sequences.

If you are interested in trying the new version of BankIt, please go [here](#). Please send any comments, questions, or suggestions to gb-admin@ncbi.nlm.nih.gov and use 'New BankIt Testing' with your BankIt ID# in the subject line. Thank you.

► **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

Internet 75%

Požadavky na každé zaslání sekvence

- kontaktní informace

Submitting Authors
File Edit

Submission Contact Authors Affiliation

First Name M.I. Last Name Sfx
Charles R Darwin

Please include country code for non-U.S. phone numbers.

Phone 01 44 171-007-1212 Fax

Email darwin@beagle.edu.uk

<< Prev Page Next >>

Submitting Authors
File Edit

Submission Contact Authors Affiliation

Institution Oxbridge University

Department Evolutionary Biology Department

Address 1859 Tennis Court Lane

City Camford

State/Province Zip/Postal Code OX1 2BH

Country United Kingdom

<< Prev Page Next Form >>

Další požadavky na zaslání sekvence

- Informace o datu zveřejnění
- Informace o relevantních publikacích
- Popis zdroje sekvence
- Vlastní sekvence
 - minimálně 50 bp
 - typ a tvar molekuly
 - anotace vlastností sekvence

Popis zdroje sekvence 1

- **organism**
nezkrácené vědecké jméno
Příklad: [organism=Drosophila melanogaster]
- **lineage**
taxonomické zařazení organismu (dle NCBI taxonomy database)
<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root>
- **molecule**
ve tvaru "DNA" nebo "RNA".
Příklad : [molecule=DNA]
- **moltype**
může nabývat následujících hodnot
Příklad : [moltype=Genomic DNA]
 - Genomic DNA
 - Genomic RNA
 - Precursor RNA
 - mRNA [cDNA]
 - Ribosomal RNA
 - Transfer RNA
 - Small nuclear RNA
 - Small cytoplasmic RNA
 - Other-Genetic
 - cRNA
 - Small nucleolar RNA
- **topology**

Popis zdroje sekvence 2

- **location**
může nabývat následujících hodnot
Příklad: [location=mitochondrion]
 - genomic
 - chloroplast
 - kinetoplast
 - mitochondrion
 - plastid
 - macronuclear
 - extrachromosomal
 - plasmid
 - cyanelle
 - proviral
 - virion
 - nucleomorph
 - apicoplast
 - leucoplast
 - proplastid
 - endogenous-virus
 - hydrogenosome
- **Genetic code**
<http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c>

Popis zdroje sekvence 3

Další popisovače ke zdroji sekvence

- acronym
- anamorph
- authority
- biotype
- biovar
- breed
- cell-line
- cell-type
- chemovar
- chromosome
- clone
- clone-lib
- collected-by
- common
- country
- cultivar
- dev-stage
- ecotype
- endogenous-virus-name
- forma
- forma-specialis
- fwd-pcr-primer-name
- fwd-pcr-primer-seq
- genotype
- group
- haplotype
- identified-by
- isolate
- isolation-source
- lab-host
- lat-lon
- map
- note
- pathovar
- plasmid-name
- plastid-name
- pop-variant
- rev-pcr-primer-name
- rev-pcr-primer-seq
- segment
- serogroup
- serotype
- serovar
- sex
- specific-host
- specimen-voucher
- strain
- sub-species
- subclone
- subgroup
- substrain
- subtype
- synonym
- teleomorph
- tissue-lib
- tissue-type
- type
- variety

Formát sekvence

- Sekvence nukleové kyseliny a kódovaných proteinů připravené ve formátu FASTA

Nucleotide Sequence:

```
>ABC-1 [organism=Saccharomyces cerevisiae][strain=ABC][clone=1]
ATTGCGTTATGGAAATTCGAAACTGCCAAATACTATGTCACCATCATTGA
TGCACCTGGACACAGAGATTTTCATCAAGAACATGATCACTGGTACTT
```

Protein Sequences:

```
>4E-I [gene=eIF4E] [protein=eukaryotic initiation factor 4E-I]
MQSDFHRMKNFANPKSMFKTSAPSTEQGRPEPPTSAAAPAEAKDVKPKEDPQETGEPAGN ...
>4E-II [gene=eIF4E] [protein=eukaryotic initiation factor 4E-II]
MVLLETEKTSAPSTEQGRPEPPTSAAAPAEAKDVKPKEDPQETGEPAGNTATTTAPAGDD ...
```

Přsrušená sekvence

```
>m_gagei [organism=Mansonia gagei] Mansonia gagei NADH dehydrogenase ...
ATGGAGCATACATATCAATATTCATGGATCATACCGTTTGTGCCACTTCCAATTCCTATTTTAATAGGAA
TTGGACTCCTACTTTTTCCGACGGCAACAAAAAATCTTCGTCGTATGTGGGCTCTTCCCAATATTTTATT
GTTAAGTATAGTTATGATTTTTTCGGTCGATCTGTCCATTCAGCAAATAAATAAAAGTTCTATCTATCAA
TATGTATGGTCTTGGACCATCAATAATGATTTTTCTTTCGAGTTTGGCTACTTTATTGATTCGCTTACCT
>?200 ← Délka přerušení
GGTATAATAACAGTATTATTAGGGGCTACTTTAGCTCTTGC
TCAAAAAGATATTAAGAGGGGTTTAGCCTATTCTACAATGTCCCAACTGGGTTATATGATGTTAGCTCTA
GGTATGGGGTCTTATCGAGCCGCTTTATTTCAATTTGATTACTCATGCTTATTTCGAAGGCATTGTTGTTTT
TAGGATCCGGATCCGTTATTCATTCCATGGAAGCTATTGTTGGATATTCTCCAGATAAAAGCCAGAATAT
GGTTTTTATGGGCGGTTTAAGAAAGCATGTGCCAATTACACAAATTGCTTTTTTTAGTGGGTACACTTTCT
CTTTGTGGTATTCCACCCCTTGCTTGTTTTTTGGTCCAAAGATGAAATTCCTTAGTGACAGCTGGTTGT
>?unk100 ← Přerušení neznámé délky
TCAATAAAACTATGGGGTAAAGAAGAACAAAAATAATTAACAGAAATTTTCGTTTATCTCCTTTATTAA
TATTAACGATGAATAATAATGAGAAGCCATATAGAATTGGTGATAATGTAAAAAAGGGGCTCTTATTAC
TATTACGAGTTTTTGGCTACAAGAAGGCTTTTTTCTTATCCTCATGAATCGGATAATACTATGCTATTTCCCT
ATGCTTATATTGGCTCTATTTACTTTTTTTGTTGGAGCCATAGCAATTCCTTTTAATCAAGAAGGACTAC
ATTTGGATATATTATCCAAATTATTA ACTCCATCTATAAATCTTTTACATCAAATTCAAATGATTTTGA
GGATTGGTATCAATTTTTAACAAATGCAACTCTTTCAGTGAGTATAGCCTGTTTCGGAATATTTACAGCA
TTCTTTTTATATAAGCCTTTTTTATTCATCTTTACAAAATTTGAACTTACTAAATTTATTTTCGAAAGGGG
GTCCTAAAAGAATTTTTTTGGATAAAATAATACTTGATATACGATTGGTCATATAATCGTGGTTACAT
```

Sekvenční příložen

- Fasta+GAP

```
>ABC-1 [organism=Saccharomyces cerevisiae][strain=ABC][clone=1]
---ATTGCGTTATGGAAATTCGAAACTGCCAAATACTATGTCACCATCAT
TGATGCACCTGGACACAGAGATTTTCATCAAGAACATGATCACTGGTACTT
>ABC-2 [organism=Saccharomyces cerevisiae][strain=ABC][clone=2]
GATATTGCTTTATGGAAATTCGAAACTGCCAAATACTATGTCACCATCAT
TGATGCACCTGGACACAGAAATTTTCATCAAGAACATGATCACTGGTACTT
>ABC-3 [organism=Saccharomyces cerevisiae][strain=ABC][clone=3]
---ATTGCTTTATGGAAATTCGAAACTGCCAAATACTATGTTA-----
TGATGCACCTGGACACAGAGATTTTCATCAAAAACATGATCACTGGTACTT
```

- PHYLIP

```
3 100
ABC-1 ---ATTGCGT TATGGAAATT CGAAACTGCC AAATACTATG TCACCATCAT
ABC-2 GATATTGCTT TATGGAAATT CGAAACTGCC AAATACTATG TCACCATCAT
ABC-3 ---ATTGCTT TATGGAAATT CGAAACTGCC AAATACTATG TTA-----

TGATGCACCT GGACACAGAG ATTTTCATCAA GAACATGATC ACTGGTACTT
TGATGCACCT GGACACAGAA ATTTTCATCAA GAACATGATC ACTGGTACTT
TGATGCACCT GGACACAGAG ATTTTCATCAA AAACATGATC ACTGGTACTT
```

```
>[organism=Saccharomyces cerevisiae][strain=ABC][clone=1]
>[organism=Saccharomyces cerevisiae][strain=ABC][clone=2]
>[organism=Saccharomyces cerevisiae][strain=ABC][clone=3]
```

Sequin – příprava zaslání sekvence

Welcome to Sequin

Misc

Sequin

Sequin Application Version 6.00
Standard Release [Oct 27 2005]

National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health

(301) 496-2475
info@ncbi.nlm.nih.gov

Database for submission GenBank EMBL DDBJ

Start New Submission

Read Existing Record

Show Help

Quit Program

Sequence Format

File

Submission type Single Sequence Segmented Sequence
 Gapped Sequence Population Study
 Phylogenetic Study Mutation Study
 Environmental Samples Batch Submission

Sequence data format FASTA (no alignment)
 Alignment (FASTA+GAP, NEXUS, PHYLIP, etc.)

Submission category Original Submission
 Third Party Annotation

<< Prev Form Next Form >>

Target Sequence eIF4E

Done

Format GenBank Mode Sequin Style Normal

CDS: eukaryotic initiation factor 4E-II

```

LOCUS       eIF4E                2881 bp    DNA     linear   INV 27-OCT-2005
DEFINITION  Drosophila melanogaster eukaryotic initiation factor 4E (eIF4E)
            gene, alternative splice products, complete cds.

ACCESSION
VERSION
KEYWORDS
SOURCE      Drosophila melanogaster (fruit fly)
            ORGANISM  Drosophila melanogaster
                    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                    Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                    Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2881)
            AUTHORS   Burnett,F.M., van der Waals,J.D. and Szent-Gyorgi,A.
            TITLE     Environmental influences on the expansion of germline tandem
                    repeats in several species of Galapagos finches
            JOURNAL    Unpublished
REFERENCE   2  (bases 1 to 2881)
            AUTHORS   Burnett,F.M., van der Waals,J.D. and Szent-Gyorgi,A.
            TITLE     Direct Submission
            JOURNAL    Submitted (27-OCT-2005) Evolutionary Biology Department, Oxbridge
                    University, 1859 Tennis Court Lane, Camford OX1 2BH, United Kingdom

FEATURES             Location/Qualifiers
     source           1..2881
                    /organism="Drosophila melanogaster"
                    /mol_type="genomic DNA"
                    /strain="Oregon R"
     gene             join(201..224,1550..1920,1986..2085,2317..2404,2466..2629)
                    /gene="eIF4E"
     CDS              join(201..224,1550..1920,1986..2085,2317..2404,2466..2629)
                    /gene="eIF4E"
                    /codon_start=1
                    /product="eukaryotic initiation factor 4E-II"
                    /translation="MVLLETEKTSAPSTEQGRPEPPTSAAAAPAEAKDVKPKEDPQETG
                    EPAGNTATTTAPAGDDAVRTEHLYKHPLMNVWTLWYLENDRSKSWEDMQNEITSFDTV
                    EDFWLSLYNHKPPSEIKLGS DYSLFKKNIRPMWEDAANKQGGRWVITLNKSSKTDLDN
                    LULDVLLCLIGEAFDHS DQICGAVINIRGKSNKISIWTDAGNNEEAAL EIGHKLRLDAL
                    RLGRMNSLQYQLHKD TMVKQGSNVKSIYTL"

```

eIF4E

File Edit Search Options Misc Annotate

Target Sequence: eIF4E Done

Format: Sequence

CDS: eukaryotic initiation factor 4E-II

Feature display: Target Numbering: Top Grid: Off

```

      10      20      30      40      50      60
      |      |      |      |      |      |
1    cggttgcttg ggttttataa catcagtcag tgacaggcat ttccagagtt gcctgttca
      70      80      90     100     110     120
61   acaatcgata gctgcctttg gccacaaaaa tcccaaactt aattaaagaa ttaaataatt
      130     140     150     160     170     180
      |      |      |      |      |      |
      aacctacgc agcttgagtg cgtaaccgat atctagtata
      210     220     230     240
      |      |      |      |      |      |
      tggtagtgt tggagacgga gaaggtaaga cgatgataga
      270     280     290     300
      |      |      |      |      |      |
      tttgcgctg agccgtggca gggaacaaca aaaacagggt
      330     340     350     360
      |      |      |      |      |      |
      atagtcgag cggaaaagag tgcagttggc gtggctacat
      390     400     410     420
      |      |      |      |      |      |
      ttttttgca caattgctta atattaattg tacttgcacg
  
```

eIF4E

File Edit Search Options Misc Annotate

Target Sequence: eIF4E Done

Format: Graphic Style: Default Filter: Default Scale: 10

eIF4E

1 1000 2000 2881

Gene: eIF4E

CDS: eukaryotic initiation factor 4E-II

CDS: eukaryotic initiation factor 4E-I

```

      210     220     230     240
      |      |      |      |      |      |
      tggtagtgt tggagacgga gaaggtaaga cgatgataga
      270     280     290     300
      |      |      |      |      |      |
      tttgcgctg agccgtggca gggaacaaca aaaacagggt
      330     340     350     360
      |      |      |      |      |      |
      atagtcgag cggaaaagag tgcagttggc gtggctacat
      390     400     410     420
      |      |      |      |      |      |
      ttttttgca caattgctta atattaattg tacttgcacg
  
```

M V V L E T E K

Coding Region File Edit

Coding Region Properties Location

Product Protein Exceptions Misc

Genetic Code Standard

Reading Frame Protein Length 248

Protein Product 4E-II

```
MVVLETEKTSAPSTEQGRPEPPTSAAAPAEAKDVI
ATTTAPAGDDAVRTEHLYKHPLMNVWTLWYLENDI
TVEDFWSLYNHKPPSEIKLGSYSLFKKNIRPMI
NKSSKTDLDNLWLDVLLCLIGEAFDHSQICGAVI
GNNEEAAL EIGHKLRDALRLGRNNSLQYQLHKDTI
```

Predict Interval Translate Product Edit

Retranslate on Accept Synchron

Accept Cancel

Coding Region File Edit

Coding Region Properties Location

General Comment Citations Cross-Refs Evidence Identifiers

Flags Partial Pseudo Evidence

Exception Explanation

Standard explanation

Gene eIF4E

Map by Overlap Cross-reference

Edit Gene Feature

Retranslate on Accept Synchron

Accept Cancel

Coding Region File Edit

Coding Region Properties Location

5' Partial 3' Partial

From	To	Strand	SeqID
201	224	Plus	eIF4E
1550	1920	Plus	eIF4E
1986	2085	Plus	eIF4E
2317	2404	Plus	eIF4E

'order' (intersperse intervals with gaps)

Retranslate on Accept Synchronize Partials

Accept Cancel

Anotace vlastní sekvence

- Kódované proteiny
 - CDS
interval
nekompletnost na N- nebo C- konci
 - gene
interval odpovídající CDS u experimentálně prokázaných genů
 - mRNA
interval obsahující 5'-UTR a 3'-UTR
- Kódované strukturní RNA

Příklady sekvencí

Sekvence mRNA nebo cDNA

- Kódující oblasti včetně iniciačního a terminačního kodonu
- Název proteinu
- Název genu
- Sekvence proteinu

Homo sapiens prolidase (PEPD) mRNA, complete cds.

FEATURES	Location/Qualifiers
source	1..1888 /organism="Homo sapiens" /chromosome="19" /map="19q12-q13.2" /cell_type="fibroblasts"
mRNA	1..1888 /gene="PEPD"
gene	1..1888 /gene="PEPD"
CDS	17..1498 /gene="PEPD" /EC_number="3.4.13.9" /note="imidodipeptidase" /product="prolidase"

Sekvence prokaryotického genu

- Kódující intervaly
- Název proteinu
- Název genu, je-li známý
- Aminokyselinová sekvence

`Escherichia coli RecA protein (recA) gene, complete cds.`

FEATURES	Location/Qualifiers
<code>source</code>	<code>1..3300</code> <code>/organism="Escherichia coli"</code> <code>/strain="K-12"</code>
<code>gene</code>	<code>783..1961</code> <code>/gene="recA"</code>
<code>CDS</code>	<code>783..1961</code> <code>/gene="recA"</code> <code>/function="DNA repair protein"</code> <code>/product="RecA protein"</code>

Sekvence eukaryotického genu

- Intervaly kódujících oblastí včetně start- a stop-kodonů a intervaly všech intronů
- Název proteinu
- Název genu, je-li známý
- Aminokyselinová sekvence

`Caenorhabditis elegans tyrosine kinase PTK-2 (ptk-2) gene, complete cds.`

FEATURES	Location/Qualifiers
<code>source</code>	<code>1..3180</code> <code>/organism="Caenorhabditis elegans"</code>
<code>gene</code>	<code>211..3011</code> <code>/gene="ptk-2"</code>
<code>mRNA</code>	<code>join(211..288,533..703,763..890,940..1024,</code> <code>1084..1380,1838..1962,2018..2099,2301..3011)</code> <code>/gene="ptk-2"</code> <code>/product="protein kinase PTK-2"</code>
<code>CDS</code>	<code>join(250..288,533..703,763..890,940..1024,</code> <code>1084..1380,1838..1962,2018..2099,2301..2456)</code> <code>/gene="ptk-2"</code> <code>/product="protein kinase PTK-2"</code>

Ribosomální RNA a vnitřní přepisované mezerníky

- Názvy jakékoli strukturní RNA (např. tRNA-Ile, 16S ribosomal RNA)
- Názvy mezerníkových oblastí (např., internal transcribed spacer 1, 16S/23S intergenic spacer)
- Nukleotidové pozice

`Saccharomyces cerevisiae 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence.`

FEATURES	Location/Qualifiers
source	1..540 /organism="Saccharomyces cerevisiae" /strain="UMD 334"
rRNA	<1..5 /product="18S ribosomal RNA"
misc_RNA	6..178 /product="internal transcribed spacer 1 "
rRNA	179..377 /product="5.8S ribosomal RNA"
misc_RNA	378..519 /product="internal transcribed spacer 2"
rRNA	520..>540 /product="28S ribosomal RNA"

Oblast promotoru

- Název proteinu nebo genu, ke kterému patří promotor a jeho 5' a 3' obklopující sekvence
- Intervaly přepisovaných a kódujících sekvencí, pokud jsou přítomné

Homo sapiens enhancer-binding protein 2 (EBP2) gene, promoter region and partial cds.

FEATURES	Location/Qualifiers
source	1..3061 /organism="Homo sapiens" /chromosome="15" /map="15q13" /cell_line="H441" /tissue_type="lung"
gene	1..>3061 /gene="EBP2"
promoter	1..2947 /gene="EBP2"
TATA_signal	2918..2923 /gene="EBP2"
mRNA	2948..>3061 /gene="EBP2" /product="enhancer-binding protein 2"
5'UTR	2948..3010 /gene="EBP2"
CDS	3011..>3061 /gene="EBP2" /product="enhancer-binding protein 2"

Transpozon nebo inzerční sekvence

Specifické jméno elementu

- Nukleotidové pozice
- Jména a intervaly kódovaných genových produktů, pokud jsou přítomny (např., transposase)
- Pozice a intervaly dalších vlastností (např. LTRs, repeat regions)

**Bacillus subtilis transposon BLT transposase (tnpA) gene,
complete cds**

```
FEATURES             Location/Qualifiers
    source             1..1221
                       /organism="Bacillus subtilis"
                       /strain="RS2"
    source             21..1127
                       /organism="Bacillus subtilis"
                       /strain="RS2"
                       /transposon="BLT"
    repeat_region      21..61
                       /rpt_type=inverted
    gene               128..1034
                       /gene="tnpA"
    CDS                128..1034
                       /gene="tnpA"
                       /product="transposase"
    repeat_region      1085..1127
                       /rpt_type=inverted
```

Oblasti repeticí

- Intervaly repetitivních sekvencí
- Rodina repeticí (např., Alu, Mer)
- Typ repetice (tandem, inverted, flanking, terminal, direct, dispersed, or other)
- Jednotka repetice (repeat unit) popis intervalů, jestliže sekvence obsahuje více než jednu repetici

Homo sapiens repeat regions

FEATURES	Location/Qualifiers
source	1..2050 /organism="Homo sapiens" /chromosome="6" /map="6q25"
repeat_region	8..126 /rpt_type=dispersed /rpt_family="B2"
repeat_region	197..344 /rpt_type="direct" /rpt_unit="197..220"
repeat_region	389..673 /rpt_family="AluSx" /rpt_type=dispersed
repeat_region	847..876 /note="microsatellite BT21" /rpt_type="tandem" /rpt_unit="ca"
repeat_region	1000..2000 /rpt_family="human endogeneous retrovirus K-10"

Klonovací vektor

- Jedinečné jméno vektoru
- Kódující intervaly, jména genů a proteinů

Cloning vector pRB223, complete sequence

FEATURES	Location/Qualifiers
source	1..4361 /organism="Cloning vector pRB223"
gene	86..1276 /gene="tet"
CDS	86..1276 /gene="tet" /product="tetracycline resistance protein"
RBS	1905..1909 /note="Shine-Dalgarno sequence"
rep_origin	2535
gene	complement(3293..4194) /gene="bla"
CDS	complement(3293..4153) /gene="bla" /product="beta-lactamase"
misc_feature	4069..4125 /note="multiple cloning site"
RBS	complement(4161..4165) /gene="bla" /note="Shine-Dalgarno sequence"
promoter	complement(4188..4194) /gene="bla"

Příklady některých dalších modifikací deskriptorů

- Title
 - Informace vyskytující se v databázi v DEFINITION LINE
- Comment
 - Poznámka k různým vlastnostem
- Technique
 - Umožňuje výběr techniky použité pro vytvoření nebo experimentální evidenci vlastností sekvence

Přehled deskriptorů pro popis vlastností sekvence

(<http://www.ncbi.nlm.nih.gov/BankIt/help.html>)

- attenuator
- C-region
- CAAT_signal
- CDS
- conflict
- D-loop
- D-segment
- enhancer
- exon
- gap
- GC_signal
- gene
- iDNA
- intron
- J_segment
- LTR
- mat_peptide
- misc_binding
- misc_difference
- misc_feature
- misc_recomb
- misc_RNA
- misc_signal
- misc_structure
- modified_base
- mRNA
- N_region
- old_sequence
- operon
- oriT
- polyA_signal
- polyA_site
- precursor_RNA
- prim_transcript
- primer_bind
- promoter
- protein_bind
- RBS
- repeat_region
- repeat_unit
- rep_origin
- rRNA
- S_region
- satellite
- scRNA
- sig_peptide
- snRNA
- snoRNA
- source
- stem_loop
- STS
- TATA_signal
- terminator
- transit_peptide
- tRNA
- unsure
- V_region
- V_segment
- variation
- 3'clip
- 3'UTR
- 5'clip
- 5'UTR