Біоінформатика.

БАЗИ ДАНИХ. МОДЕЛЬ NCBI

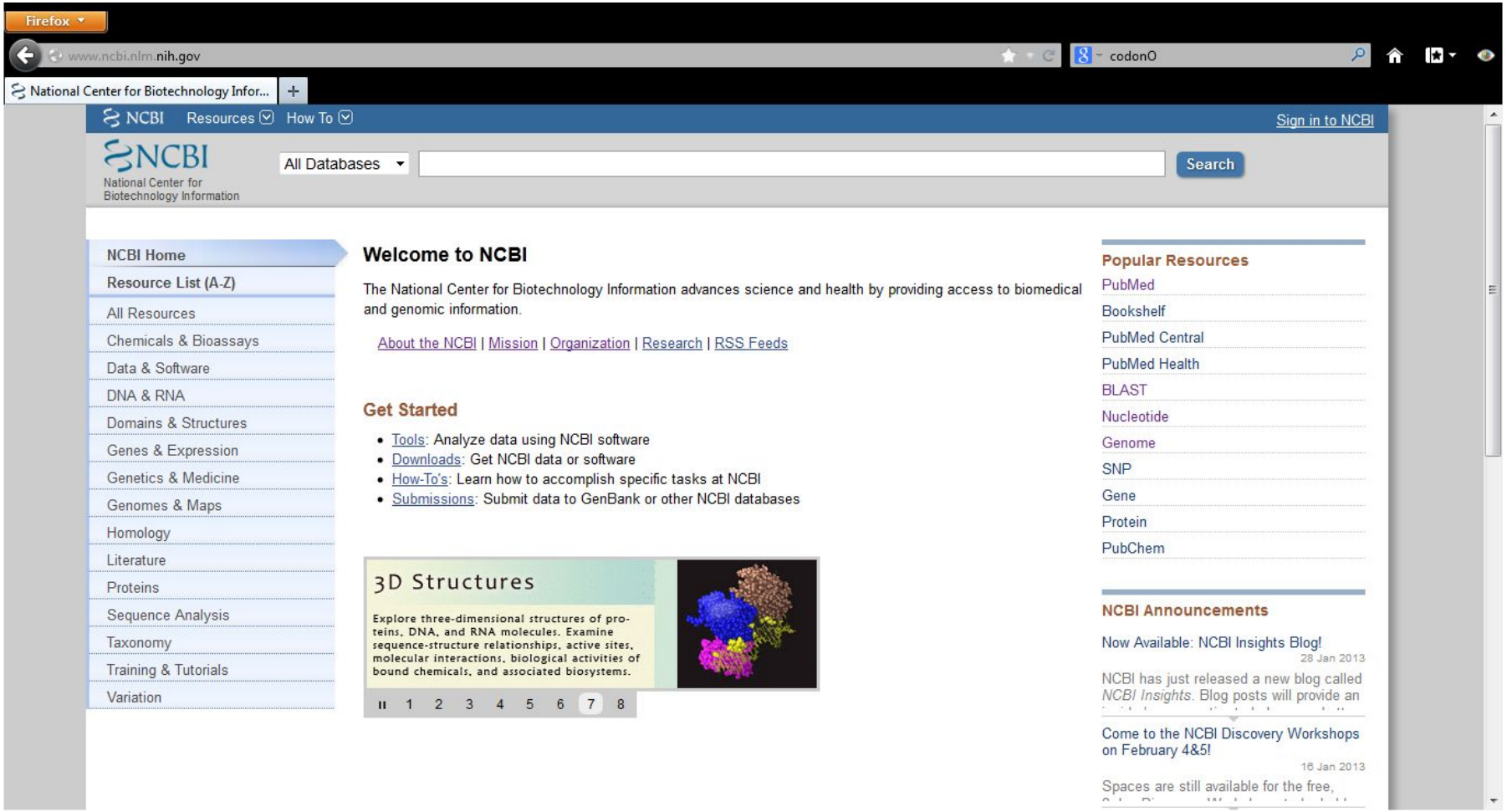
Lab.1

Центральні питання

- Де знайти літературу про ген/білок, що цікавить
- Де можна знайти послідовність гена/генома, що цікавить?
- Що таке депонування НАП?

National Center for Biotechnology Information

- www.ncbi.nlm.nih.gov



NCBI

- Фізична локація – м. Бетезда, штат Меріленд (США)
- Подвійна функція – сховище даних і сервер для аналізу

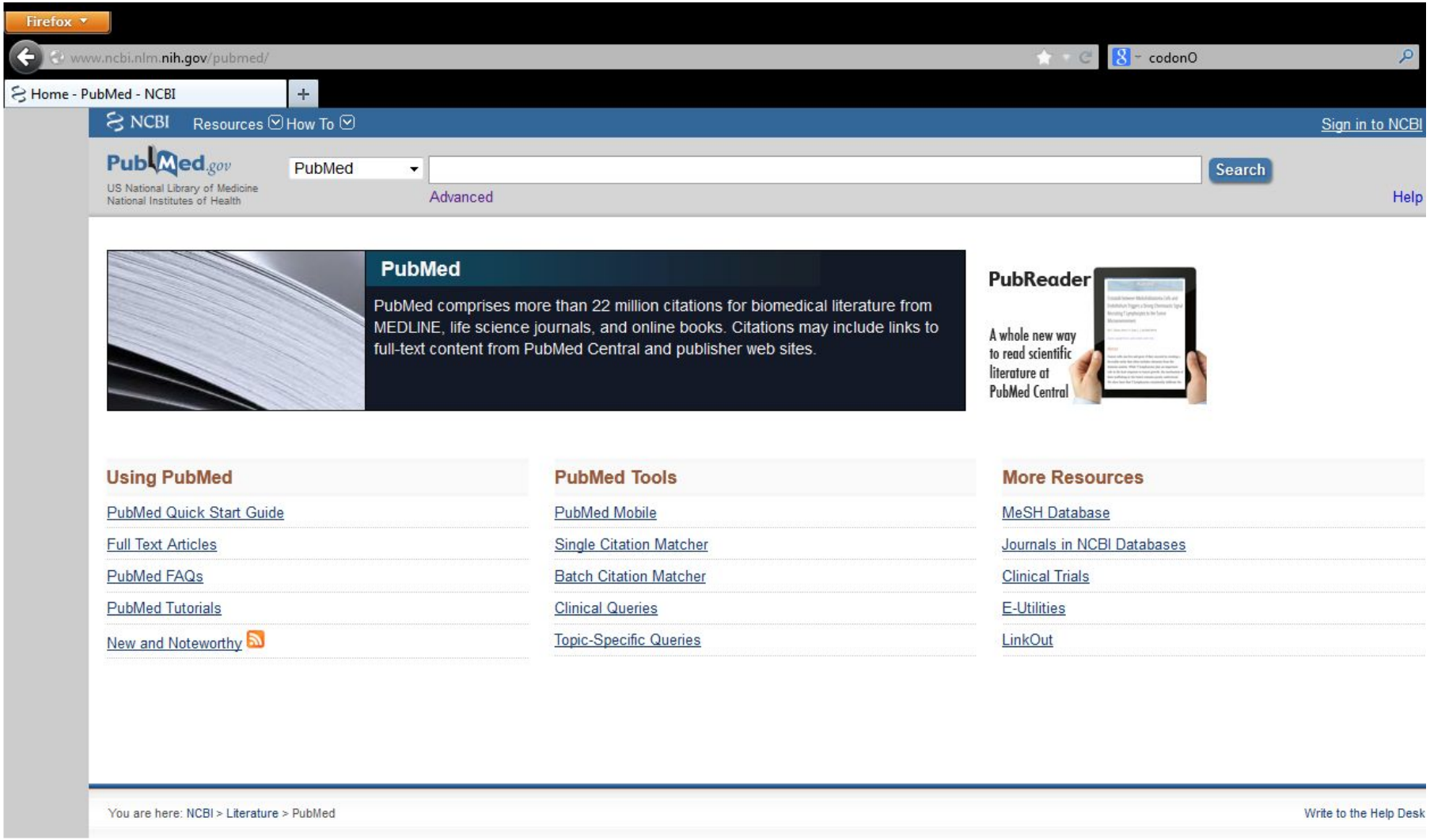
Типи даних

Біомедичні дані
PubMed
PubChem
Taxonomy
Chemicals and assays
Proteins
…..

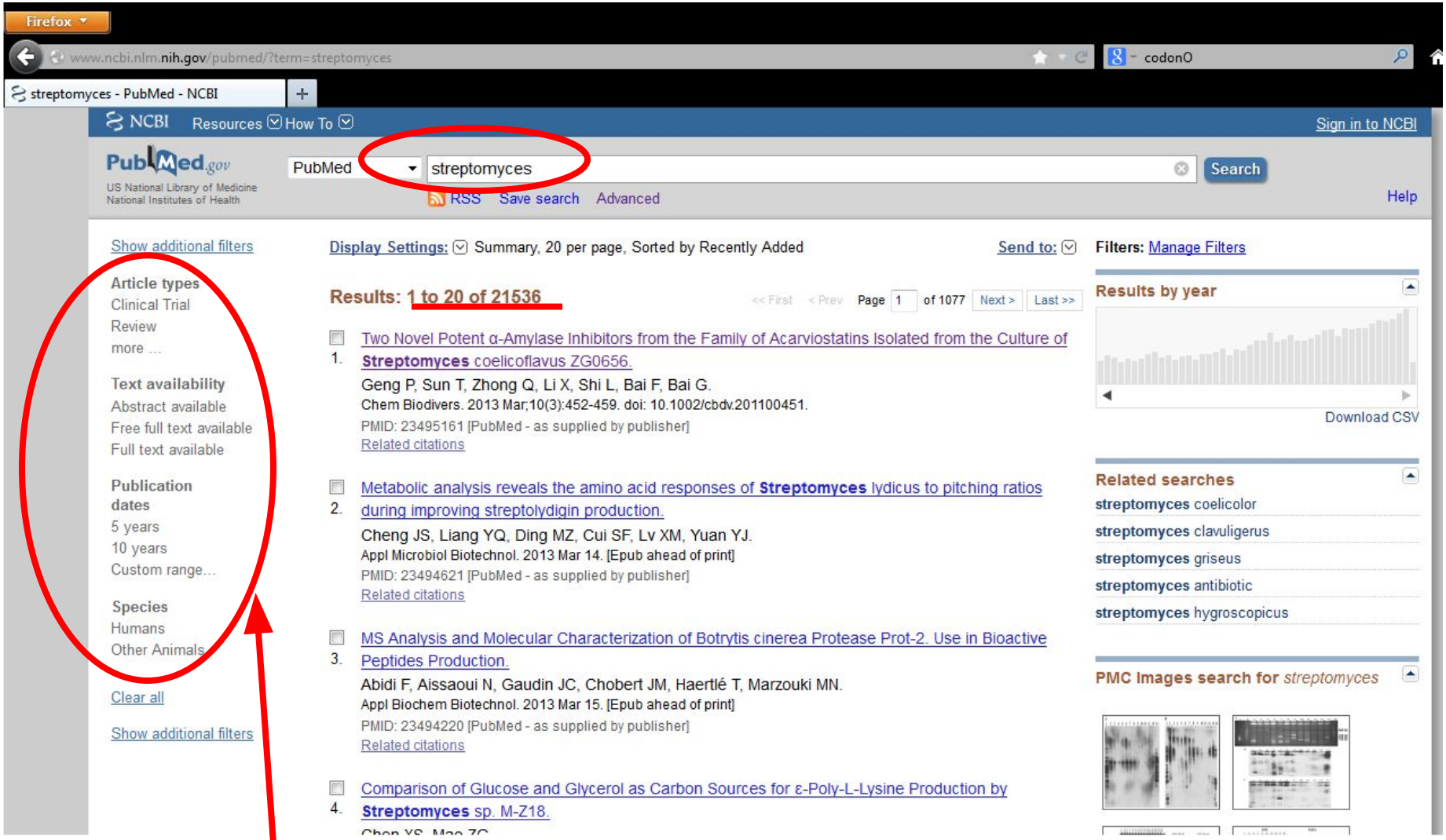
Геномні дані
Genome
Variations
SNP
OMIM
Genes and expression
…..

PubMed

- www.ncbi.nlm.nih.gov/pubmed



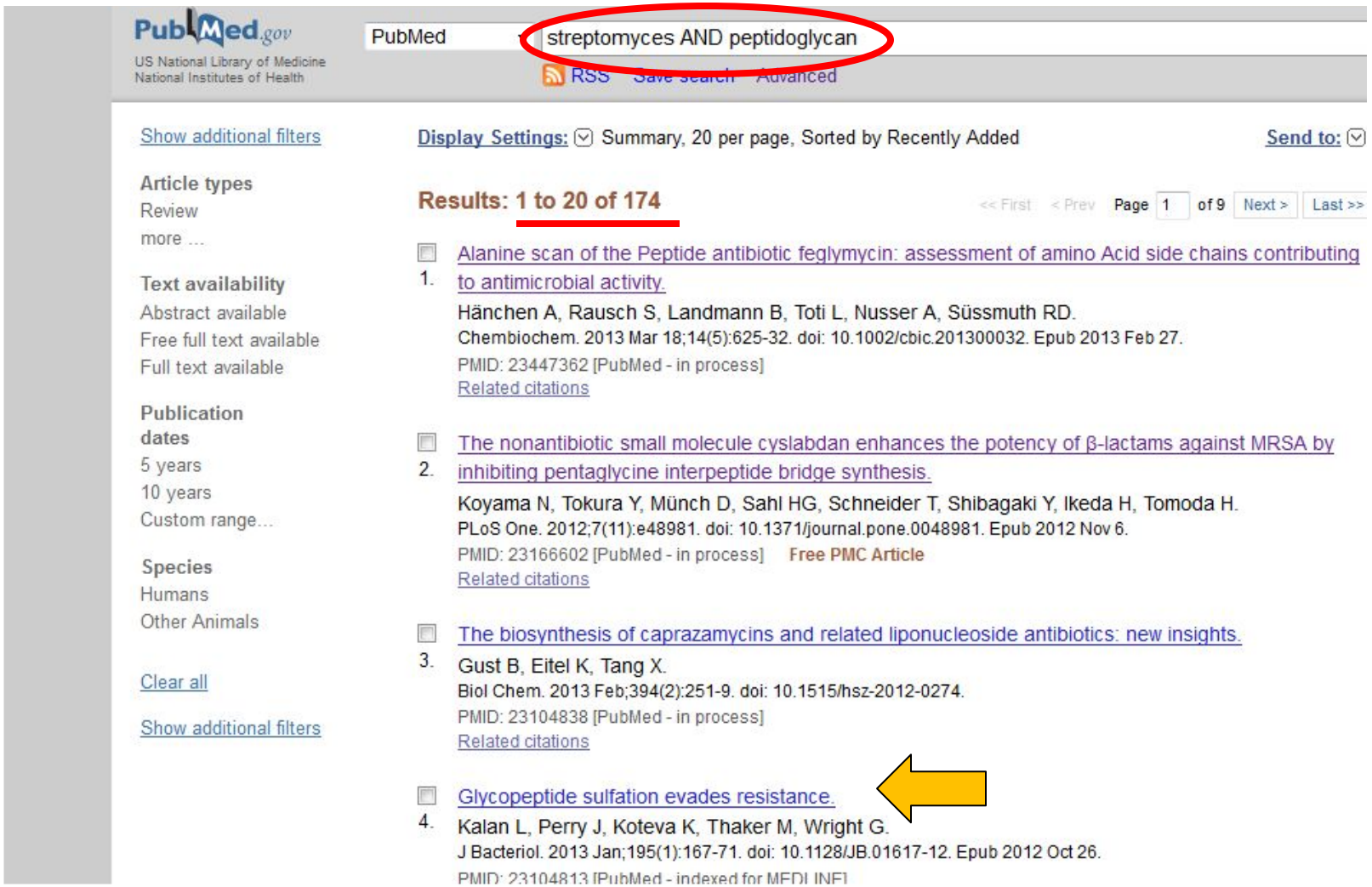
PubMed



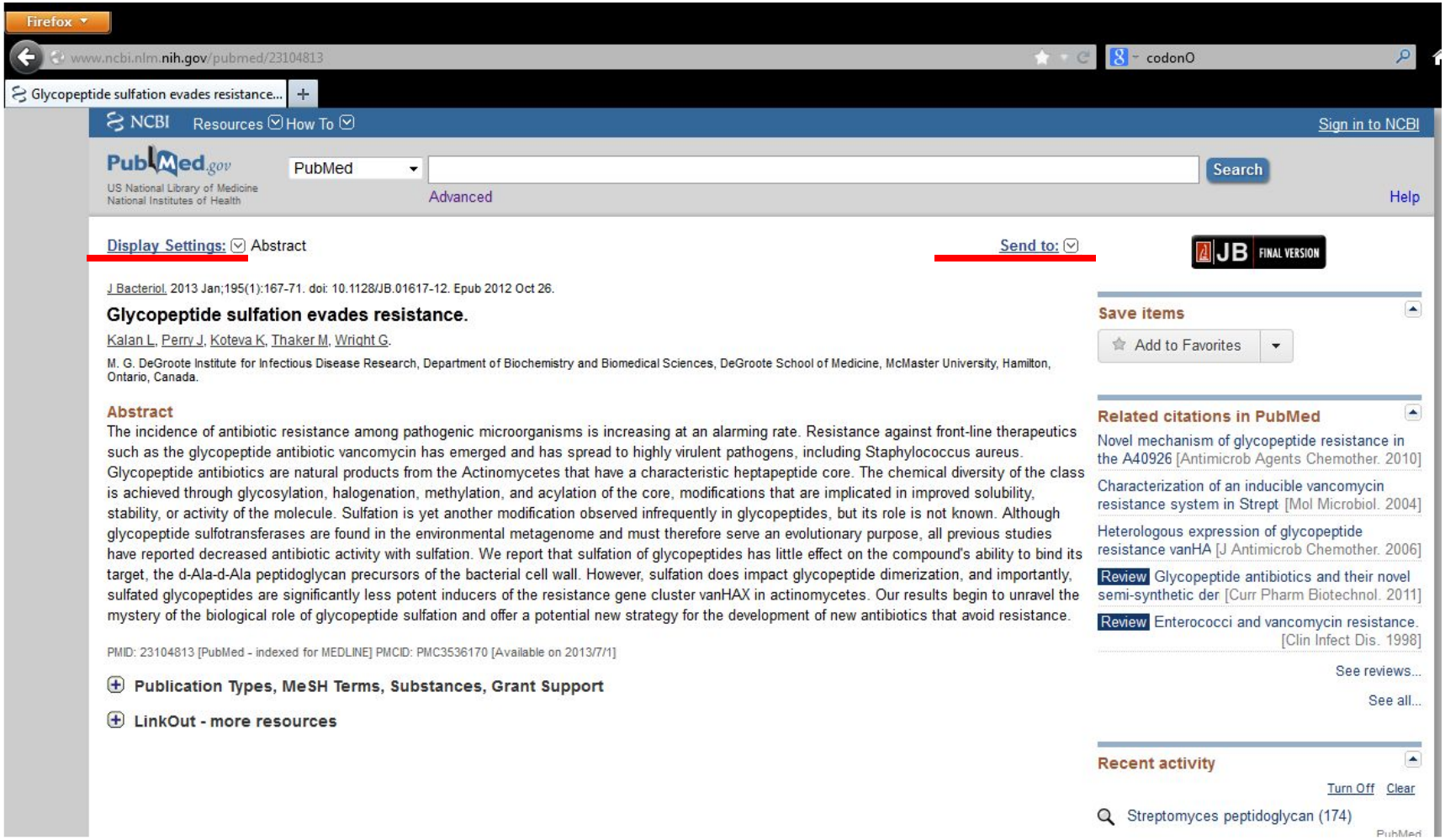
Можна звузити коло пошуку до статей, що відповідають певним вимогам (напр., шукати тільки огляди - review)

PubMed

Логічні оператори: AND, OR, *.
Streptomyces; strepto*. * = будь-що



PubMed


Firefox
www.ncbi.nlm.nih.gov/pubmed/23104813
codonO
Glycopeptide sulfation evades resistance...
NCBI Resources How To
Sign in to NCBI
PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
Search
Advanced
Help
Display Settings: Abstract
Send to:
JB FINAL VERSION
J Bacteriol. 2013 Jan;195(1):167-71. doi: 10.1128/JB.01617-12. Epub 2012 Oct 26.
Glycopeptide sulfation evades resistance.
Kalan L, Perry J, Koteva K, Thaker M, Wright G.
M. G. DeGroote Institute for Infectious Disease Research, Department of Biochemistry and Biomedical Sciences, DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada.
Abstract
The incidence of antibiotic resistance among pathogenic microorganisms is increasing at an alarming rate. Resistance against front-line therapeutics such as the glycopeptide antibiotic vancomycin has emerged and has spread to highly virulent pathogens, including Staphylococcus aureus. Glycopeptide antibiotics are natural products from the Actinomycetes that have a characteristic heptapeptide core. The chemical diversity of the class is achieved through glycosylation, halogenation, methylation, and acylation of the core, modifications that are implicated in improved solubility, stability, or activity of the molecule. Sulfation is yet another modification observed infrequently in glycopeptides, but its role is not known. Although glycopeptide sulfotransferases are found in the environmental metagenome and must therefore serve an evolutionary purpose, all previous studies have reported decreased antibiotic activity with sulfation. We report that sulfation of glycopeptides has little effect on the compound's ability to bind its target, the d-Ala-d-Ala peptidoglycan precursors of the bacterial cell wall. However, sulfation does impact glycopeptide dimerization, and importantly, sulfated glycopeptides are significantly less potent inducers of the resistance gene cluster vanHAX in actinomycetes. Our results begin to unravel the mystery of the biological role of glycopeptide sulfation and offer a potential new strategy for the development of new antibiotics that avoid resistance.
PMID: 23104813 [PubMed - indexed for MEDLINE] PMCID: PMC3536170 [Available on 2013/7/1]
Publication Types, MeSH Terms, Substances, Grant Support
LinkOut - more resources
Save items
Add to Favorites
Related citations in PubMed
Novel mechanism of glycopeptide resistance in the A40926 [Antimicrob Agents Chemother. 2010]
Characterization of an inducible vancomycin resistance system in Strept [Mol Microbiol. 2004]
Heterologous expression of glycopeptide resistance vanHA [J Antimicrob Chemother. 2006]
Review Glycopeptide antibiotics and their novel semi-synthetic der [Curr Pharm Biotechnol. 2011]
Review Enterococci and vancomycin resistance. [Clin Infect Dis. 1998]
See reviews...
See all...
Recent activity
Turn Off Clear
Streptomyces peptidoglycan (174)
PubMed

PubMed

Скориставшись лінками, що підкреслено на попередньому слайді, можна отримати бібліографічний опис статті:

1: Kalan L, Perry J, Koteva K, Thaker M, Wright G. Glycopeptide sulfation evades resistance. J Bacteriol. 2013 Jan;195(1):167-71. doi: 10.1128/JB.01617-12. Epub 2012 Oct 26. PubMed PMID: 23104813; PubMed Central PMCID: PMC3536170.

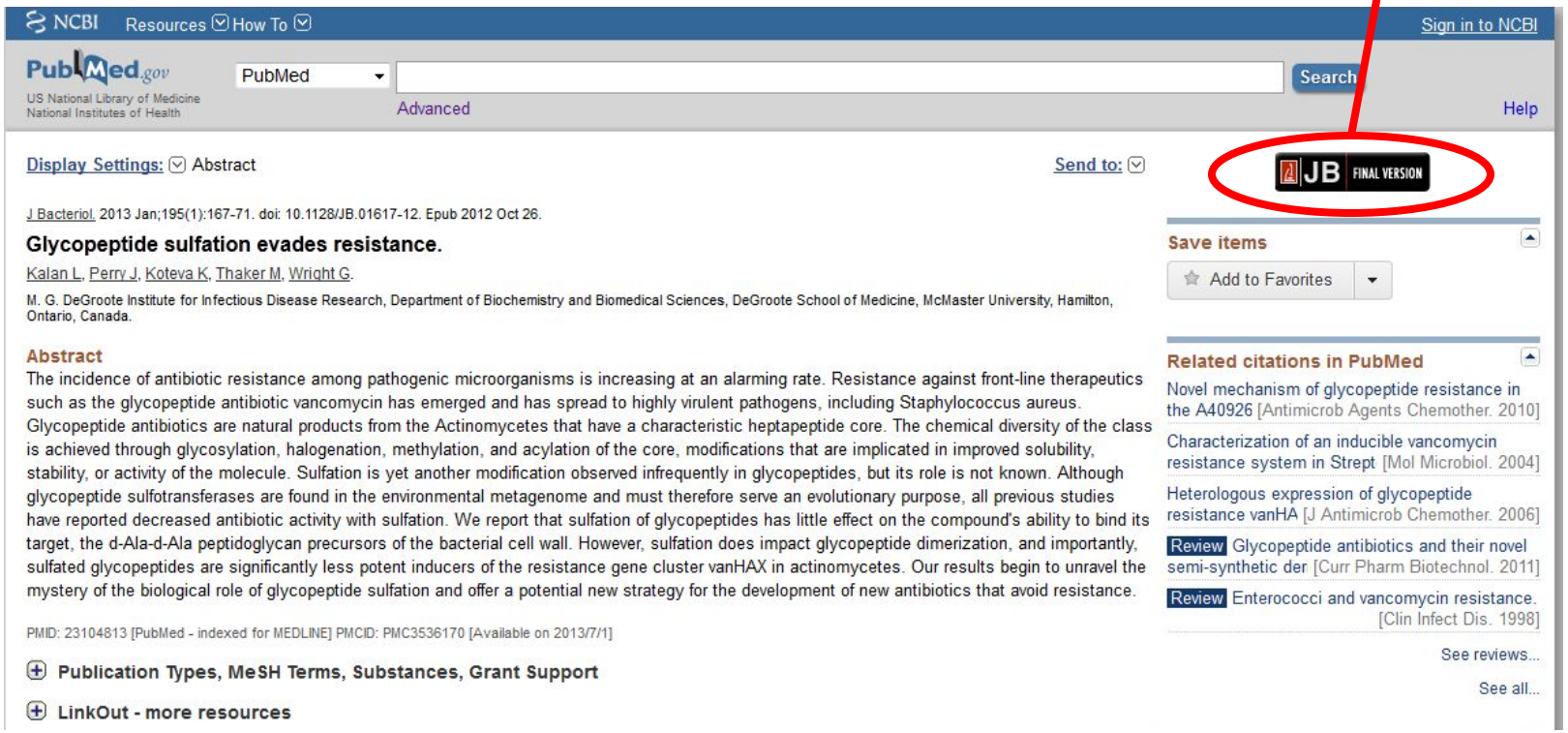
Ідентифікатори статті як цифрового об'єкта (ЦО):

- **DOI (Digital Object Identification) – унікальна комбінація арабських цифр і латинських літер – жоден інший ЦО такої не має. Видавець звертається до www.doi.org, що видає первинний номер (10.1128/JB), який далі продовжується у міру появи нових ЦО. DOI можна присвоїти частинам статті (рисункам)**
- **PMID – присвоюється автоматично кожному посиланню, що потрапляє у базу PubMed**
- **PMCID – номери доступу до повних текстів статей, зібраних із доступних джерел фахівцями NCBI, і поміщених в PubMed Central**

PubMed

Доступ до повних текстів статей

Сайт видавця



PubMed

Деякі видання у галузі хімії не реферуються в PubMed

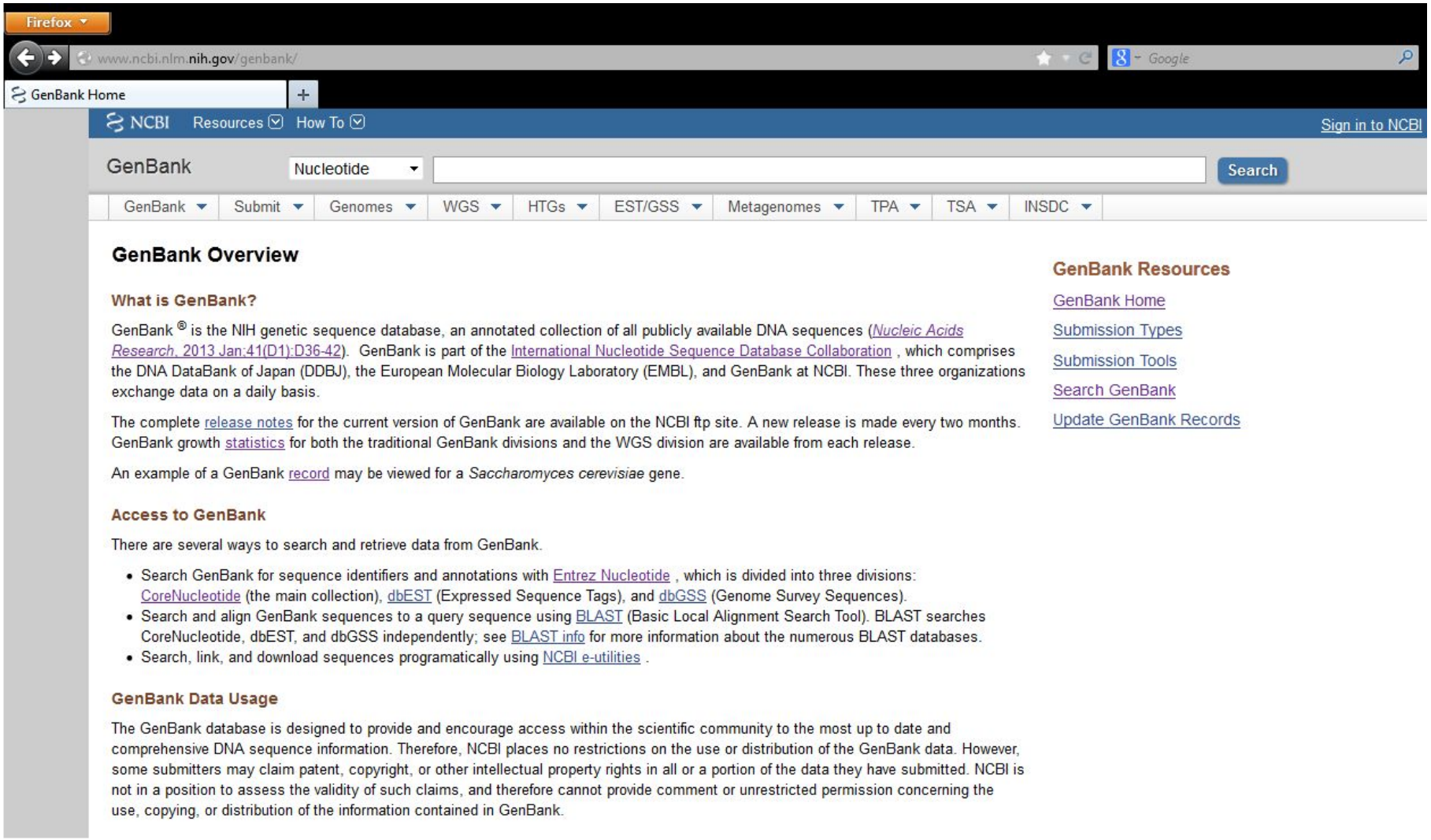
Видавничий дім Elsevier

- **Журнали Open Access**
- **NIH вимагає відкривати зміст усіх статей, де описано досліди, що ними фінансувались – через певний період часу (зразу, півроку, рік)**
- **Запит до автора статті (може переслати персональну копію для особистого вжитку)**
- **Пошук через Google**
- **Запит через наукові соцмережі (LinkedIn, ResearchGate)**

GenBank

- Ця база є частиною міжнародного консорціуму INSDC (International Nucleotide Sequence Databases Collaboration; http://www.insdc.org/), куди входять European Nucleotide Archive (ENA) і DNA Data Base of Japan (DDBJ).
- GenBank – підрозділ NCBI. www.ncbi.nlm.nih.gov/genbank
- Т.зв. первинна, або архівна, база даних – метою є збереження нуклеотидних послідовностей (і продуктів їхньої концептуальної трансляції), що подаються (депонуються) дослідниками з усього світу. Зміст поданої інформації загалом не перевіряється – це відповідальність авторів депонування
- GenBank, ENA і DDBJ щоденно обмінюються інформацією, і зміст їхніх баз ідентичний

GenBank



Анотовані фрагменти геномів; анотовані геноми; неанотовані фрагменти і геноми; чернетки (драфти геномів), метагеномні дані….

GenBank

Формат FASTA

Шрифт `Courier New`

```
>PM13114
TTAAGAAGCAATAGGTGTCCATAGTTCACCTAATTTATAGTCCCACT
CTTTTGGTTTAATGGGAATGCGCCCCGCATTGGGGGTAAAGGTCATT
TCACTAAACAGTAGTGTGGTTTCTGACATTAAAAAATCGACCCGACA
ATAGGCAAAACCGTTAGCTAATTTGTTGGCTAATGTGAGCATATTGT
CAAATTGTAGCGGTTTTTCAATATATTCTGGGGTATTGGGATCTTCT
AAAGTAAAAGGCTGTAATTGCCATTGTGTATCATAAACATTGATA
```

Рядок анотації може займати кілька рядків, містити пробіли

```
>WemR
MKKLKKYLTRKKKENYIIFSIYYFIKVTSSIFISDSLYRKYIFKRKY
KLNLKKPTSFNEKIHYRILNDHNPIYTKLADKLLVRDYVREKIGEKY
LIKLINHYNTPSEINFNTLPKSFVLKCNHDVGSVMIINDKSKINEKA
IKKKLKIALKNNIYYQNREWHYKNIKPKIICEELINIFPHNKKNYPE
DYKIHCFNGIPRYIELQFSRFSHDRRINIYDFNWNLQPFLMGYKNTN
ESIEKPKKLQEIYNISKTLSADFDYCRVDFYITP
```

GenBank

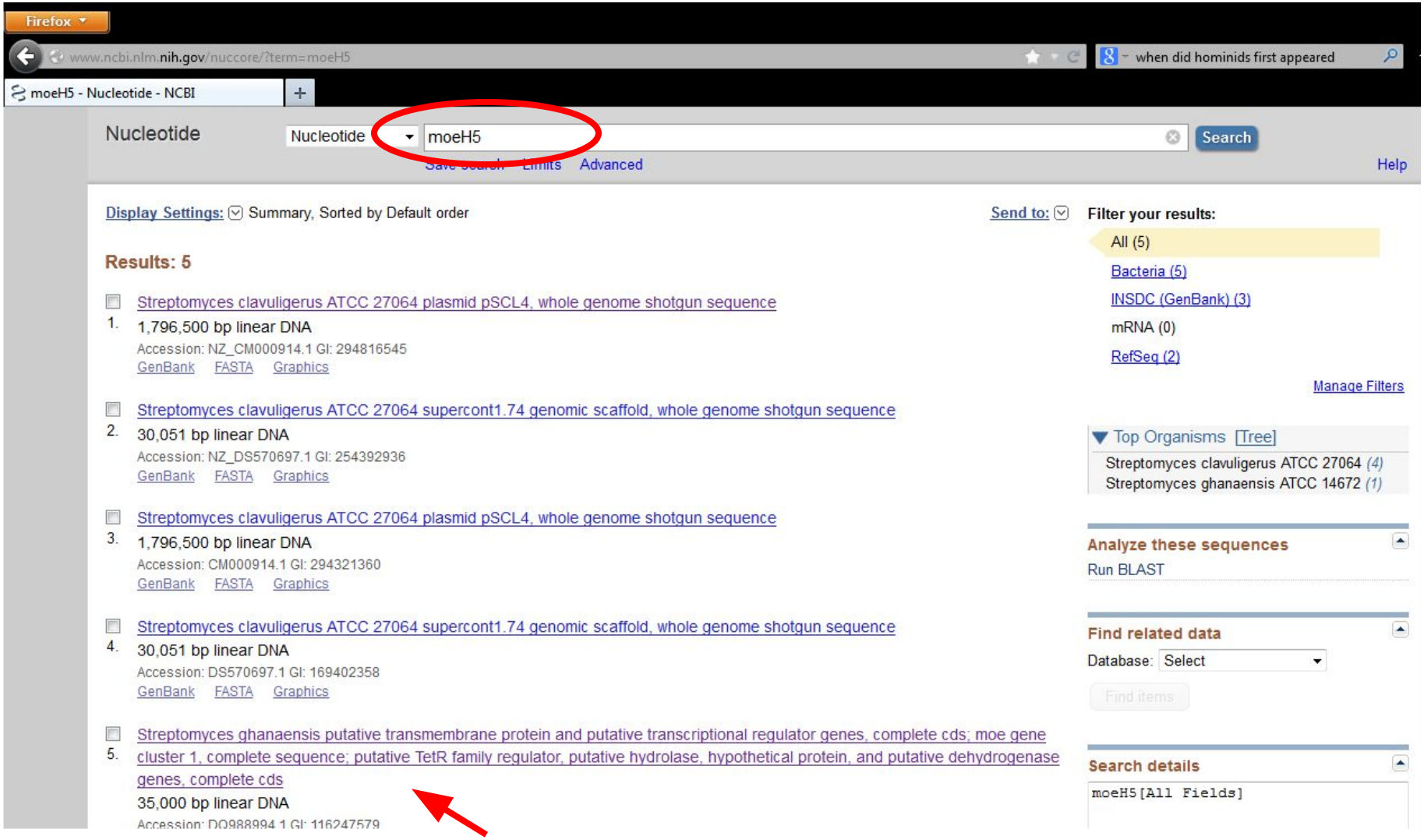
Gene identification (gi) – унікальний номер, що присвоюється кожній послідовності при поданні у GenBank. Після двокрапки – координати у геномі (у п.н.)

Анотація

Locus name

```
>gi|197283915:3424858-3425733|PMI3114| Proteus
mirabilis HI4320 chromosome, complete genome
TTAAGAAGCAATAGGTGTCCATAGTTCACCTAATTTATAGTCCCAC
TCTTTTGGTTTAATGGGAATGCGCCCCGCATTGGGGGTAAAGGTCA
TTTCACTAAACAGTAGTGTGGTTTCTGACATTAAAAAATCGACCCG
ACAATAGGCAAAACCGTTAGCTAATTTGTTGGCTAATGTGAGCATA
TTGTCAAATTGTAGCGGTTTTTCAATATATTCTGGGGTATTGGGAT
CTTCTAAAGTAAAAGGCTGTAATTGCCATTGTGTATCATAAACATT
GATA
```

GenBank


Firefox
www.ncbi.nlm.nih.gov/nuccore/?term=moeH5
when did hominids first appeared
moeH5 - Nucleotide - NCBI
Nucleotide
Nucleotide
moeH5
Search
Limits
Advanced
Help
Display Settings: Summary, Sorted by Default order
Send to:
Results: 5
1. Streptomyces clavuligerus ATCC 27064 plasmid pSCL4, whole genome shotgun sequence
1,796,500 bp linear DNA
Accession: NZ_CM000914.1 GI: 294816545
GenBank FASTA Graphics
2. Streptomyces clavuligerus ATCC 27064 supercont1.74 genomic scaffold, whole genome shotgun sequence
30,051 bp linear DNA
Accession: NZ_DS570697.1 GI: 254392936
GenBank FASTA Graphics
3. Streptomyces clavuligerus ATCC 27064 plasmid pSCL4, whole genome shotgun sequence
1,796,500 bp linear DNA
Accession: CM000914.1 GI: 294321360
GenBank FASTA Graphics
4. Streptomyces clavuligerus ATCC 27064 supercont1.74 genomic scaffold, whole genome shotgun sequence
30,051 bp linear DNA
Accession: DS570697.1 GI: 169402358
GenBank FASTA Graphics
5. Streptomyces ghanaensis putative transmembrane protein and putative transcriptional regulator genes, complete cds; moe gene cluster 1, complete sequence; putative TetR family regulator, putative hydrolase, hypothetical protein, and putative dehydrogenase genes, complete cds
35,000 bp linear DNA
Accession: DQ988994.1 GI: 116247579
Filter your results:
All (5)
Bacteria (5)
INSDC (GenBank) (3)
mRNA (0)
RefSeq (2)
Manage Filters
Top Organisms [Tree]
Streptomyces clavuligerus ATCC 27064 (4)
Streptomyces ghanaensis ATCC 14672 (1)
Analyze these sequences
Run BLAST
Find related data
Database: Select
Find items
Search details
moeH5[All Fields]

GenBank

Структура GenBank FlatFile (GBFF) - флетфайл

GenBank

HEADER

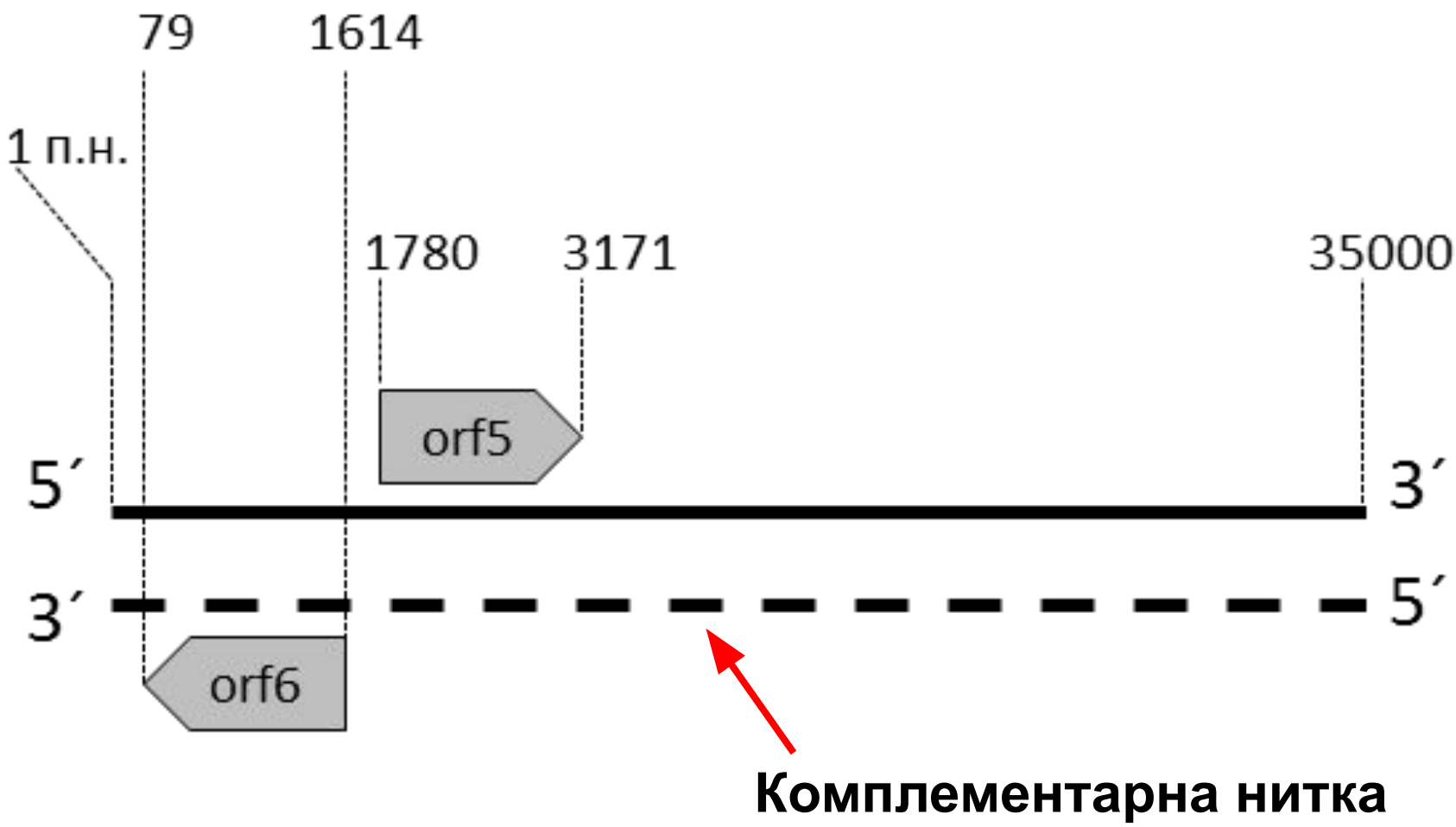
```
LOCUS       DQ988994               35000 bp    DNA     linear   BCT 04-APR-2007
DEFINITION  Streptomyces ghanaensis putative transmembrane protein and putative
            transcriptional regulator genes, complete cds; moe gene cluster 1,
            complete sequence; putative TetR family regulator, putative
            hydrolase, hypothetical protein, and putative dehydrogenase genes,
            complete cds.
ACCESSION   DQ988994
VERSION     DQ988994.1  GI:116247579
KEYWORDS    .
SOURCE      Streptomyces ghanaensis ATCC 14672
  ORGANISM  Streptomyces ghanaensis ATCC 14672
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 35000)
  AUTHORS   Ostash,B., Saghatelian,A. and Walker,S.
  TITLE     A streamlined metabolic pathway for the biosynthesis of moenomycin
            A
  JOURNAL   Chem. Biol. 14 (3), 257-267 (2007)
   PUBMED   17379141
REFERENCE   2  (bases 1 to 35000)
  AUTHORS   Ostash,B.O. and Walker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2006) Microbiology and Molecular Genetics,
            Harvard Medical School, 200 Longwood Ave., Boston, MA 02115, USA
```

GenBank

FEATURES

```
FEATURES             Location/Qualifiers
     source          1..35000
                     /organism="Streptomyces ghanaensis ATCC 14672"
                     /mol_type="genomic DNA"
                     /strain="ATCC14672"
                     /db_xref="ATCC:14672"
                     /db_xref="taxon:566461"
     CDS             complement(79..1614)
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transmembrane protein"
                     /protein_id="ABJ90150.1"
                     /db_xref="GI:116247580"
                     /translation="MKTAVTDIETYGVEQIPDEDRTARPLDLFRLAFGGANTFATCVL
                     GTFPILFGLSFWQGLAATLLGLVAGALLLAPMALFGPTNGTSNSVSSSAHLGVHGRVV
                     GSFLALLTAIAFFSISVWSSGDALVGGAHRLLGVPESTVTYALAYAVFAGLVLVVCIY
                     GFRFMLLVNKIAVVAASALFVLGAFAFAGDFDPGYAGTFASTADPLFWPSFIGSALIV
                     LSNPVSFGAFLGDWSRYIPAATPRRRVMGAAFLAQIATLLPFVFGLATASIIATRAGA
                     YLDPDAPNYVGGLLAIAPGWYFLPLCLIALIGGLSTGTTALYGTGLDFCSVFTRFSRV
                     QATIFIGVLSIVFIFAGRFWLNLTQSISTFATLIITCTAPWMVVMILGYVTRRGWYDP
                     EALQVFNQRRTGGRYWFAHGWNWRGLATWLVSAAVALLFTNLPGQFVGPLGDLADGMD
                     VSLPVGLALAAVLYLVLLTLFPEPRGVYGPDGPRFVRTSDAEVLPISGGPREAPAQPA
                     VVG"
```

GenBank


79
1614
1 п.н.
1780
3171
35000
orf5
5´
3´
3´
5´
orf6
Комплементарна нитка

GenBank

FEATURES

```
gene            13627..15168
                /gene="moeH5"
CDS             13627..15168
                /gene="moeH5"
                /note="putative amidotransferase subunit 2"
                /codon_start=1
                /transl_table=11
                /product="MoeH5"
                /protein_id="ABJ90159.1"
                /db_xref="GI:116247589"
                /translation="MTVRRPAASAPRVLLTAGPDGVRVEGDGEARLGHPLTGDHLDPG
                PPAEGVFAGWRWDGERLVARNDRYGVCPLFYRAGGGSLALSPDPLALLPEDGPVELDH
                DALAVFLRTGFFLAEDTAFAQVRALPPAATLTWDTGGLRLRSDGPPRPGAAAMTEAQA
                VDGFVDLFRASVARRLPGEPYDLPLSGGRDSRHILLELCRRGAPPRRCVSGAKFPPDP
                GADARVAAALAGRLGLPHTVVPRPRSQFRAELAALPAQGMTTLDGAWTQPVLAHLRRH
                SRISYDGLGGGELVQNPSVEFIRANPYDPADLPGLADRLLAASRTGPHVEHLLSPRTN
                ALWSRQAARRRLVTELARHADSASPLSSFFFWNRTRRSISAAPFALGDGRVLTHTPYL
                DHALFDHLASVPHRFLVDGTFHDRALHRAFPEHADLGFASSVPQRHGPVLVAHRLAYL
                LRFLAHATVVEPGWWRGPDRFLQRLLAAGRGPGAPQRVSRLQPLALYLLQLEDLAVRR
                ARRRP"
```

GenBank

ORIGIN

```
ORIGIN
        1 cgcgcccctc cggaggctgt ccggaagggc gcggcacggg tgggggcggg tcggccgggt
       61 tccgccggcc gggtcgggtc agccgacgac ggccggctgt gcgggcgcct cccgcgggcc
      121 gccgctgatc ggcaggacct cggcgtccga ggtccgcacg aagcggggac cgtccggccc
      181 gtacaccccg cgcggctccg ggaacagggt gagcaggacg aggtagagca ccgccgccag

     ..................

    34441 tacggccgtg gcgaattggt ggcgggccgg ggcgaaggtc caggaggcgt acaggactcc
    34501 gcgccgggca agggcgacca gggtgcagaa accggtgtct ccgtgcacga actgccgggt
    34561 gccgtcgacg gggtcgacta tccagaccgg cgcctcgccc cgaaccgcct cgtacgacgt
    34621 cgggttggcg tgcacggcct cctcgcccac cacgaccgag ccgggcagca gggcggtgag
    34681 cgcctccgtg aggtacagct ccgccttgcg gtcggcgtcg gtcacgaggt cgtgcgggcc
    34741 gctcttcagg tccacctcgt gttcggcgag ccggcgccag cgcggcatga tctcctgcgc
    34801 ggcggccttg cggacggctt cctccacgtc gacggcgtgc cggtcgagaa actcttcgat
    34861 ggtttcgttg tccttgatca tgcctccatg agaccacgcc cggccgacgt tccccaccgt
    34921 cccggtgcac tgcgggggga atcggcatga atatggggtg ccggaccacg ggcgggtggg
    34981 ctcagcggcc caccgcgtag
//
```

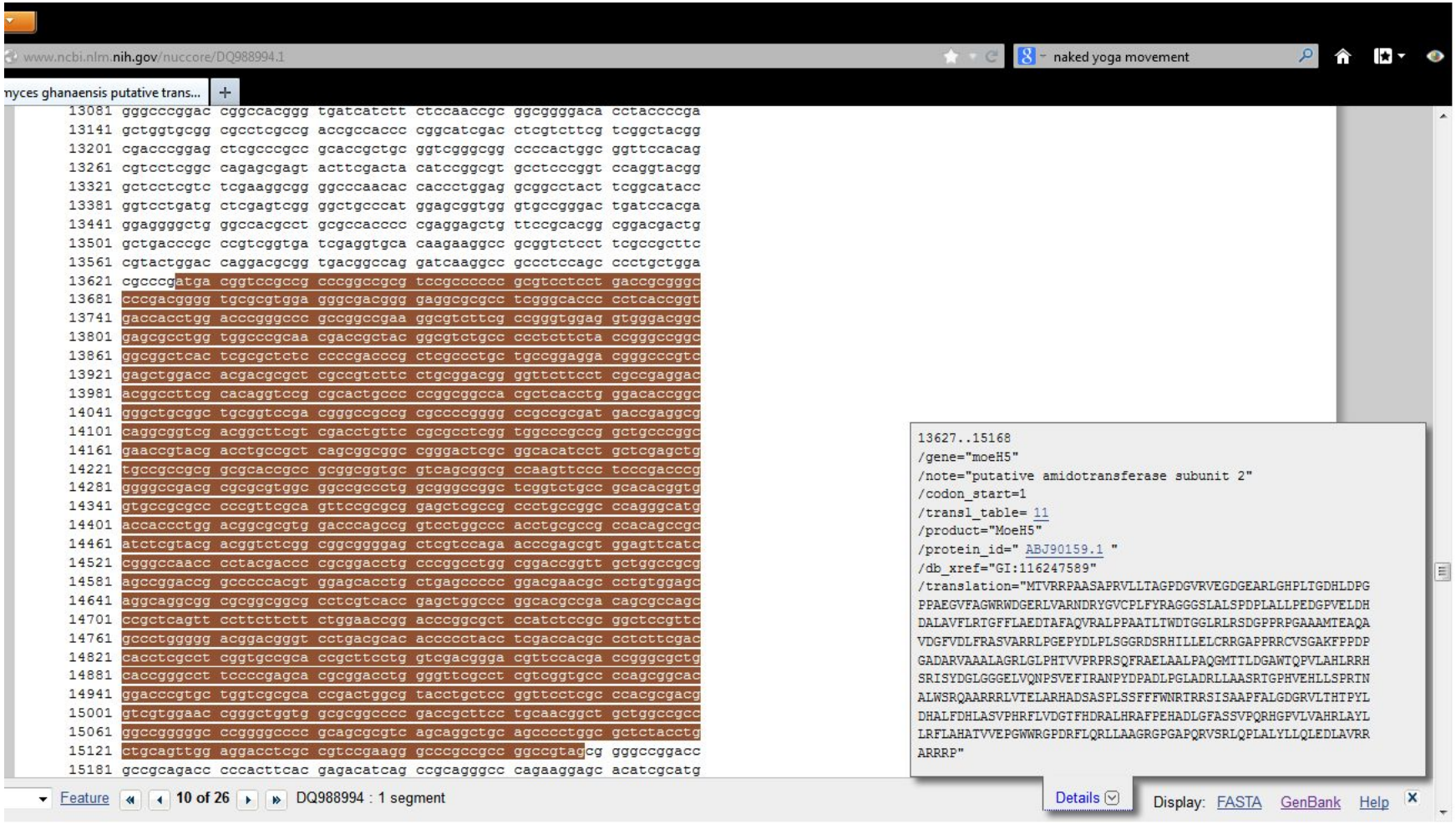
GenBank

FEATURES

```
gene            13627..15168
                /gene="moeH5"
CDS             13627..15168
                /gene="moeH5"
                /note="putative amidotransferase subunit 2"
                /codon_start=1
                /transl_table=11
                /product="MoeH5"
                /protein_id="ABJ90159.1"
                /db_xref="GI:116247589"
                /translation="MTVRRPAASAPRVLLTAGPDGVRVEGDGEARLGHPLTGDHLDPG
                PPAEGVFAGWRWDGERLVARNDRYGVCPLFYRAGGGSLALSPDPLALLPEDGPVELDH
                DALAVFLRTGFFLAEDTAFAQVRALPPAATLTWDTGGLRLRSDGPPRPGAAAMTEAQA
                VDGFVDLFRASVARRLPGEPYDLPLSGGRDSRHILLELCRRGAPPRRCVSGAKFPPDP
                GADARVAAALAGRLGLPHTVVPRPRSQFRAELAALPAQGMTTLDGAWTQPVLAHLRRH
                SRISYDGLGGGELVQNPSVEFIRANPYDPADLPGLADRLLAASRTGPHVEHLLSPRTN
                ALWSRQAARRRLVTELARHADSASPLSSFFFWNRTRRSISAAPFALGDGRVLTHTPYL
                DHALFDHLASVPHRFLVDGTFHDRALHRAFPEHADLGFASSVPQRHGPVLVAHRLAYL
                LRFLAHATVVEPGWWRGPDRFLQRLLAAGRGPGAPQRVSRLQPLALYLLQLEDLAVRR
                ARRRP"
```

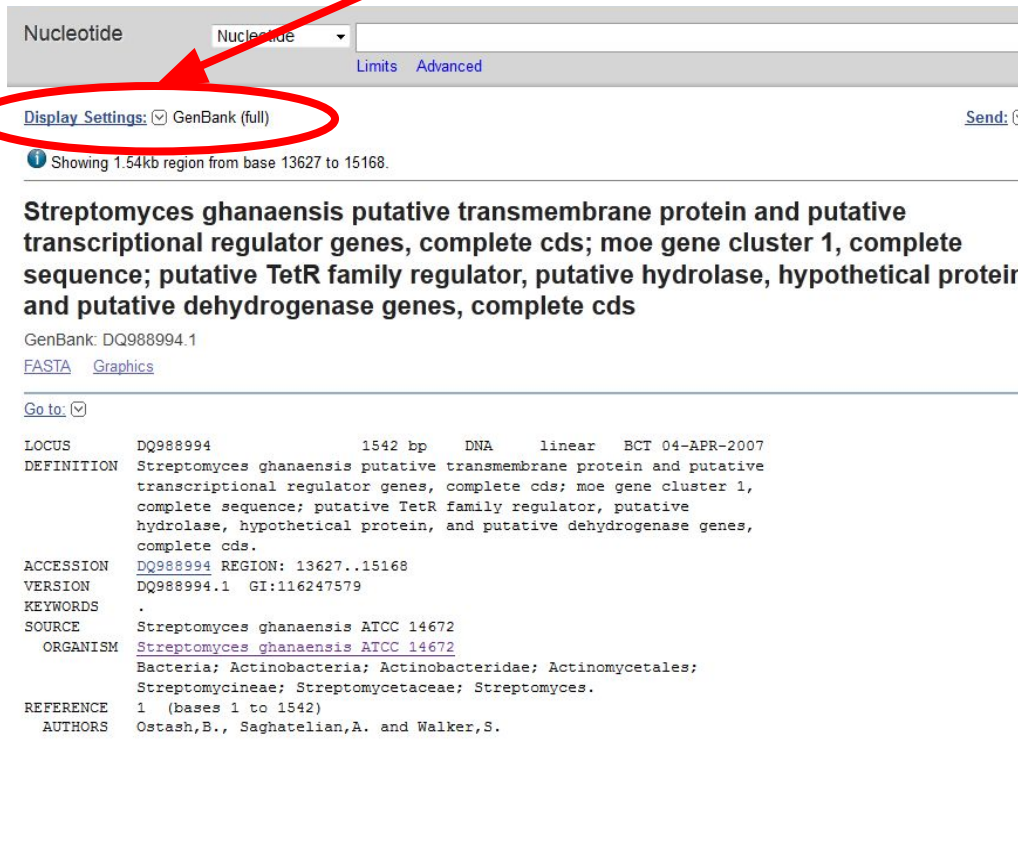
GenBank

moeH5



GenBank

GBFF для *moeH5* - гена



GenBank

FASTA-файл для *moeH5* - гена

Display Settings: FASTA Send:

Showing 1.54kb region from base 13627 to 15168.

Streptomyces ghanaensis putative transmembrane protein and putative transcriptional regulator genes, complete cds; moe gene cluster 1, complete sequence; putative TetR family regulator, putative hydrolase, hypothetical protein, and putative dehydrogenase genes, complete cds

GenBank: DQ988994.1

GenBank Graphics

```
>gb|DQ988994.1|:13627-15168 Streptomyces ghanaensis putative transmembrane
protein and putative transcriptional regulator genes, complete cds; moe gene
cluster 1, complete sequence; putative TetR family regulator, putative
hydrolase, hypothetical protein, and putative dehydrogenase genes, complete cds
ATGACGGTCCGCCGCCCGGCCGCGTCCGCCCCCCGCGTCCTCCTGACCGCGGGCCCCGACGGGGTGCGCG
TGGAGGGCGACGGGGAGGCGCGCCTCGGGCACCCCCTCACCGGTGACCACCTGGACCCGGGCCCGCCGGC
CGAAGGCGTCTTCGCCGGGTGGAGGTGGGACGGCGAGCGCCTGGTGGCCCGCAACGACCGCTACGGCGTC
TGCCCCCTCTTCTACCGGGCCGGCGGCGGCTCACTCGCGCTCTCCCCCGACCCGCTCGCCCTGCTGCCGG
AGGACGGGCCCGTCGAGCTGGACCACGACGCGCTCGCCGTCTTCCTGCGGACGGGGTTCTTCCTCGCCGA
GGACACGGCCTTCGCACAGGTCCGCGCACTGCCCCCGGCGGCCACGCTCACCTGGGACACCGGCGGGCTG
CGGCTGCGGTCCGACGGGCCGCCGCGCCCCGGGGCCGCCGCGATGACCGAGGCGCAGGCGGTCGACGGCT
TCGTCGACCTGTTCCGCGCCTCGGTGGCCCGCCGGCTGCCCGGCGAACCGTACGACCTGCCGCTCAGCGG
CGGCCGGGACTCGCGGCACATCCTGCTCGAGCTGTGCCGCCGCGGCGCACCGCCGCGGCGGTGCGTCAGC
GGCGCCAAGTTCCCTCCCGACCCGGGGGCCGACGCGCGCGTGGCGGCCGCCCTGGCGGGCCGGCTCGGTC
TGCCGCACACGGTGGTGCCGCGCCCCCGTTCGCAGTTCCGCGCGGAGCTCGCCGCCCTGCCGGCCCAGGG
CATGACCACCCTGGACGGCGCGTGGACCCAGCCGGTCCTGGCCCACCTGCGCCGCCACAGCCGCATCTCG
TACGACGGTCTCGGCGGCGGGGAGCTCGTCCAGAACCCGAGCGTGGAGTTCATCCGGGCCAACCCCTACG
ACCCCGCGGACCTGCCCGGCCTGGCGGACCGGTTGCTGGCCGCGAGCCGGACCGGCCCCCACGTGGAGCA
CCTGCTGAGCCCCCGGACGAACGCCCTGTGGAGCAGGCAGGCGGCGCGGCGGCGCCTCGTCACCGAGCTG
GCCCGGCACGCCGACAGCGCCAGCCCGCTCAGTTCCTTCTTCTTCTGGAACCGGACCCGGCGCTCCATCT
CCGCGGCTCCGTTCGCCCTGGGGGACGGACGGGTCCTGACGCACACCCCCTACCTCGACCACGCCCTCTT
CGACCACCTCGCCTCGGTGCCGCACCGCTTCCTGGTCGACGGGACGTTCCACGACCGGGCGCTGCACCGG
GCCTTCCCCGAGCACGCGGACCTGGGGTTCGCCTCGTCGGTGCCCCAGCGGCACGGACCCGTGCTGGTCG
```

GenBank

FASTA-файл для *moeH5* - гена

Display Settings: FASTA Send:

Showing 1.54kb region from base 13627 to 15168.

Streptomyces ghanaensis putative transmembrane protein and putative transcriptional regulator genes, complete cds; moe gene cluster 1, complete sequence; putative TetR family regulator, putative hydrolase, hypothetical protein, and putative dehydrogenase genes, complete cds

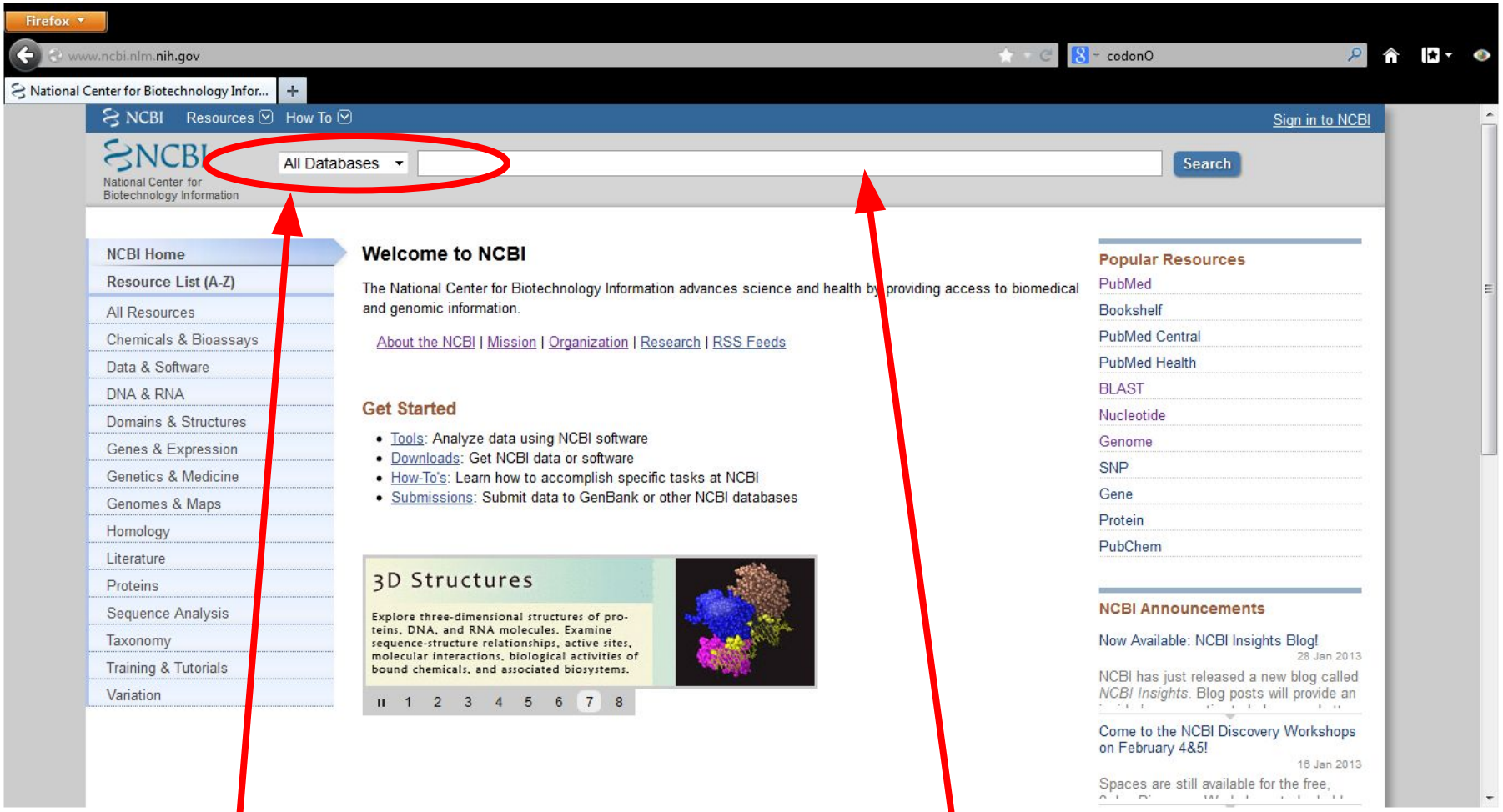
GenBank: DQ988994.1

GenBank Graphics

```
>gb|DQ988994.1|:13627-15168 Streptomyces ghanaensis putative transmembrane
protein and putative transcriptional regulator genes, complete cds; moe gene
cluster 1, complete sequence; putative TetR family regulator, putative
hydrolase, hypothetical protein, and putative dehydrogenase genes, complete cds
ATGACGGTCCGCCGCCCGGCCGCGTCCGCCCCCCGCGTCCTCCTGACCGCGGGCCCCGACGGGGTGCGCG
TGGAGGGCGACGGGGAGGCGCGCCTCGGGCACCCCCTCACCGGTGACCACCTGGACCCGGGCCCGCCGGC
CGAAGGCGTCTTCGCCGGGTGGAGGTGGGACGGCGAGCGCCTGGTGGCCCGCAACGACCGCTACGGCGTC
TGCCCCCTCTTCTACCGGGCCGGCGGCGGCTCACTCGCGCTCTCCCCCGACCCGCTCGCCCTGCTGCCGG
AGGACGGGCCCGTCGAGCTGGACCACGACGCGCTCGCCGTCTTCCTGCGGACGGGGTTCTTCCTCGCCGA
GGACACGGCCTTCGCACAGGTCCGCGCACTGCCCCCGGCGGCCACGCTCACCTGGGACACCGGCGGGCTG
CGGCTGCGGTCCGACGGGCCGCCGCGCCCCGGGGCCGCCGCGATGACCGAGGCGCAGGCGGTCGACGGCT
TCGTCGACCTGTTCCGCGCCTCGGTGGCCCGCCGGCTGCCCGGCGAACCGTACGACCTGCCGCTCAGCGG
CGGCCGGGACTCGCGGCACATCCTGCTCGAGCTGTGCCGCCGCGGCGCACCGCCGCGGCGGTGCGTCAGC
GGCGCCAAGTTCCCTCCCGACCCGGGGGCCGACGCGCGCGTGGCGGCCGCCCTGGCGGGCCGGCTCGGTC
TGCCGCACACGGTGGTGCCGCGCCCCCGTTCGCAGTTCCGCGCGGAGCTCGCCGCCCTGCCGGCCCAGGG
CATGACCACCCTGGACGGCGCGTGGACCCAGCCGGTCCTGGCCCACCTGCGCCGCCACAGCCGCATCTCG
TACGACGGTCTCGGCGGCGGGGAGCTCGTCCAGAACCCGAGCGTGGAGTTCATCCGGGCCAACCCCTACG
ACCCCGCGGACCTGCCCGGCCTGGCGGACCGGTTGCTGGCCGCGAGCCGGACCGGCCCCCACGTGGAGCA
CCTGCTGAGCCCCCGGACGAACGCCCTGTGGAGCAGGCAGGCGGCGCGGCGGCGCCTCGTCACCGAGCTG
GCCCGGCACGCCGACAGCGCCAGCCCGCTCAGTTCCTTCTTCTTCTGGAACCGGACCCGGCGCTCCATCT
CCGCGGCTCCGTTCGCCCTGGGGGACGGACGGGTCCTGACGCACACCCCCTACCTCGACCACGCCCTCTT
CGACCACCTCGCCTCGGTGCCGCACCGCTTCCTGGTCGACGGGACGTTCCACGACCGGGCGCTGCACCGG
GCCTTCCCCGAGCACGCGGACCTGGGGTTCGCCTCGTCGGTGCCCCAGCGGCACGGACCCGTGCTGGTCG
```

Прямий шлях до MoeH5

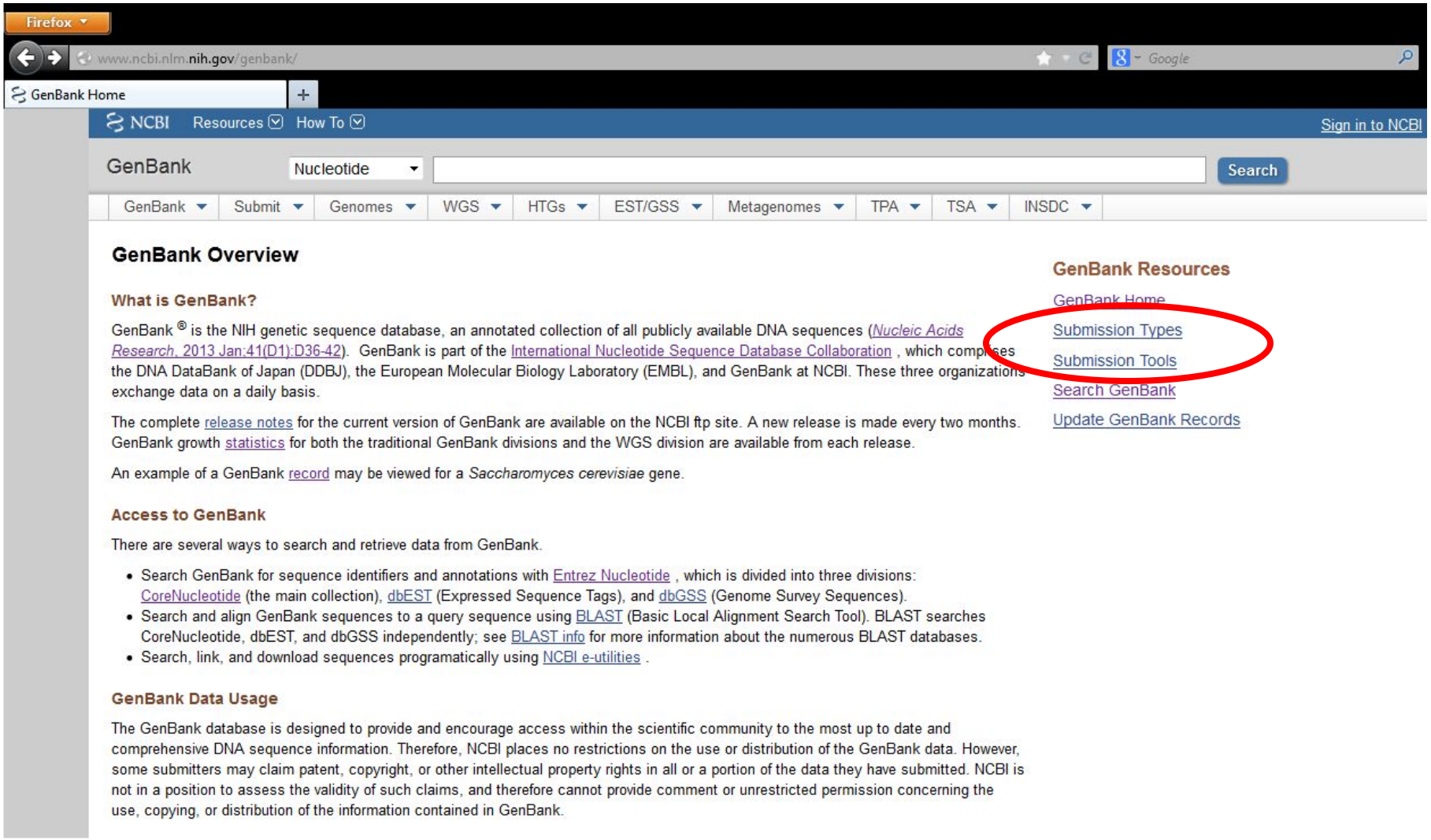
Стартова сторінка NCBI



Вибрати Protein

Задати назву білка (MoeH5)

Депонування НАП



Депонування НАП

- Нуклеотидні послідовності – у базу, що входить INSDC - GenBank, ENA, DDBJ
- Білкові – SWISSPROT
- GenBank має різні знаряддя для депонування:
  - BankIt – для подання відносно коротких і простих (безінтронних) послідовностей
  - SequIn – подання складних послідовностей
  - Та інші (тут не розглядаються)
- Подання геномів – окрема історія (див. далі)
- При поданні нуклеотидних послідовностей потрібно знати мінімум інформації про неї, зокрема локалізацію в ній (координати) orf
- У результаті подання послідовності присвоюється номер доступу в GenBank

Депонування НАП

Мінімум інформації про послідовність

- Розмір (у п.н.)
- Локалізація orf (якщо є)
- Походження послідовності (що секвенували – хромосома, плазміда тощо; з якого організму)

1

167

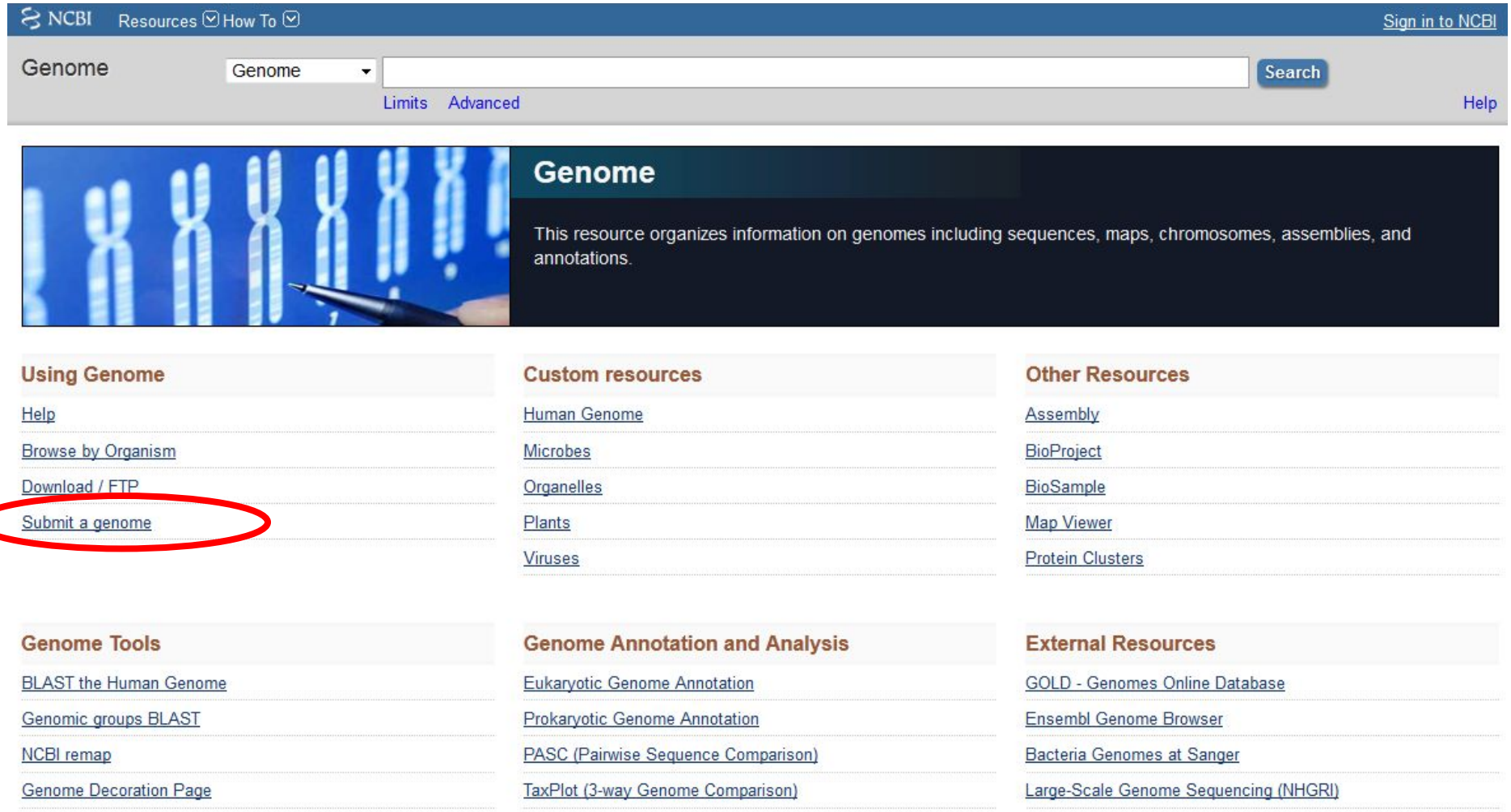
CTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCGCAGGAAGCGAGAGCCCCCGGGCCTCCTATCCTGGGCGGGCGCGAA
CGGCTTCGGCCCGACGCGAACGACCGTAGGGGCAAAGGCCATCGGGGGATCGGTTCAGCATGGGCAAGGGCAAGGACAGCCAGGGGTACCCGGCGATGGAGGGCATA
CGGCCGCTCGTGACCGGTGACCCGGCGCGGATCGGGCCCTATCGGCTACTGGGACGGCTCGGTGCGGGCGGGATGGGCCGGGTGTACCTGGCCCGTTCGGCGGGCGG
GCGCACCGTCGCCGTGAAGGTGGTGCACGAGGAGCACATCGCGAACGGGGAGTTCCGGGCCCGCTTCCGTCGGGAGATCGAGGCCGCCCGCCGCGTCGGCGGGCGGT
ACACCGCTCCCGTGCTCGACGCCGACGCCGACGCCGAACGGCCCTGGGTGGCCACCGGCTACGTCCCCGGCCCCTCCCTCGAACAGGCCGTACGGGAGCACGGGCCG
CTGCCCGCCGCCTCGGTCAACGCCCTGGCCGAGGGGCTGTTGAGGGCCCTGCGCGGCATTCACGCCGCCGGGATCGTCCACCGGGACCTCAAGCCGTCCAACGTCCT
GCTCACCGTCGACGGCCCCCGCGTCATCGACTTCGGCATCGCCCGCGCGCTCCAGGTCTCCGTGGAGTCCCTGCTGACCAGTACGGGCATGGTCATCGGCTCTCCCG
GGTTCATGGCGCCCGAGCAGATCCTCGGCGAGGAGACCGGAGCGGGGGCGGACGTCTTCTCCCTCGGCTGTGTCCTGATGTACGCGGCCACGGGGCGGCTGCCGTTC
GGCGCCGGTGCGAGCAACCAGCACGCGGTGATGTTCCGGATCGTCCAGTCCCCGCCGGACCTCGACGCGGTGGAGGACGCCTCCCTGCGCGAGCTGATCGAACGCTG
CCTGACCAAAGCCGCCGCCGAACGCCCGGGAGTGGACGAGCTGGTGGGGTGCCTGGCGCCTGACCGGTCCTCGGACGCCTTGCGTGGTGCCTGGCTGCCGCCGGTGC
TCCTGGCCCGGCTGGCGCAGCAGTCGGCGCTGCTGCTCGACGCGGATGTTCCGGAGACCTCGGACGTACCGGCGCCTGTGCGGGCGGGGGGCGCCGAGGAGCCTGCT
TCTTCGGTTGTTCCTGAGGCTCCCGTTGTTCCTGAGGCTCCCGTGCCCGAGCCGCGTGACCTCGGTACGGTCGATCTGAGGCCGGCACCGGCGAAGGACGACGAGGC
GGGCAGCCTGTCACCGGACCCGGCCCCGGACCCGGACCCGGACTCTTCCCCGGCATCGGCATCGGCATCGGGCCCGGCCCCGGCCTTGTCCCCTGGTCCCGGAGGGG
GCTGGTGCCCGCCGGTGGCGGGCGGACCGGCCAGGGAAGTGGGCGTAGGCGCTCGTGCGTGGGTCGTCGCGGCAGTCGTCGTCGCCGTGCTGGCTGCCGGGGGCACC
ACCGCCTTCCTCAACCGCGGCCCGGGTGGCGCCGACCCCGGGGGCGGAGAGGCTGCCGCGCCGCCCGCCGCCAGTGACGGACCGTCCGGCGCCCCGTCCTCCCCGGC
CGGGGAGGACGACTCCAAGGGCAGCGAGAAGGACAAGGGCAAGGACAAGGACAACGGCAAGGAAGACGGGAAGGACAAGGAAGGTACAGAAGAGGAGCAGGGCGGAG
AGGACGGCGGCGACTCCGGCGGCGGGCAGAGCGACGGGGCCGGTTCCGGCGAGGGCGCGTCCTCGGGCGGAGGCGAGGGGGCCGGAGCTTCCGGCGGAGGCGGCACA
CAGGACGGGTCCGGCTCGTCCGGTTCGGGGGGCTCCGGCGGTGCCGCGAAGCCGCCCGCCCAGGACCCCGCCCCGGACGGCCGGGTGCCGCAGCAGTTCGTCGGGAC
CTGGTCCATCGCGTCCCAGTACGACGCCCTCCAGCCGCACACCGTGGTCATCCGGCGGGTGTCGCCCGGGCAGTCGGCGGTGACCCTCATCGCCGACGTCCAGGGGT
CGGGGCACTGCGAGTACACGGCGAAGCTCAGCTCCGTGGCGGACGGCGGGAACCGGATCAACGTCGGTACCGCCGTGGTGGACAAGGCCCGTTCCGGCGGGGTGTGC
CGCGACACCGACCCCTCGTTCTTCACCGTCGCCGGCTCCGGCATCCTGCACGACGTCGGGCCCGCCCACGGCAGCGGCTATCGCTACAACCGCGCCTGAGCGGGCCC
GTCACTCCGGAATCGACGGTCCGGAACCGGCCGACGTCCGACGCCCG

2239

2294

Genome

www.ncbi.nlm.nih.gov/genome



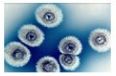
Genome

Genome | Genome | streptomyces coelicolor | Search

Save search | Limits | Advanced | Help

Display Settings: Overview | Send to:

Organism Overview ; Genome Project Report ; Genome Annotation Report ; Plasmid Annotation Report



Streptomyces coelicolor

Well-studied antiobiotic-producing bacterium

Lineage: Bacteria[4068]; Actinobacteria[697]; Actinobacteria[697]; Actinobacteridae[646]; Actinomycetales[623]; Streptomycineae[67]; Streptomycetaceae[67]; Streptomyces[65]; Streptomyces coelicolor[1]

Streptomyces. These bacteria are widely distributed in nature, especially in the soil. The characteristic earthy smell of freshly plowed soil is actually attributed to the aromatic terpenoid geosmin produced by species of *Streptomyces*. There are currently 364 known species of this genus, many of which are the most important industrial More...

Representatives

1. **Community selected** : Streptomyces coelicolor
2. **Calculated** : Streptomyces coelicolor A3(2)

***Streptomyces coelicolor* strain A3(2) M145**. This strain is a derivative of the laboratory strain A3(2) lacking its two plasmids SCP1 and SCP2 which were sequenced separately after being isolated from the original strain A3(2).

Human Pathogen: no

Type	Name	RefSeq	INSDC	Size (Mb)	GC%	Protein	rRNA	tRNA	Other RNA	Gene	Pseudogene
Chr	-	NC_003888.3	AL645882.2	8.67	72.1	7,768	18	65	3	7,911	56
Plsm	SCP2	NC_003904.1	AL645771.1	0.031317	72.1	34	-	-	-	34	-
Plsm	SCP1	NC_003903.1	AL589148.1	0.356023	69.1	351	-	-	-	355	4

Biological Properties

- Morphology
  - Gram : Positive
  - Shape : Tailed
  - Shape : Filamentous
  - Motility : Yes
- Environment
  - OxygenReq : Aerobic
  - OptimumTemperature : 25-35
  - TemperatureRange : Mesophilic
  - Habitat : Multiple
- Phenotype
  - Disease : None

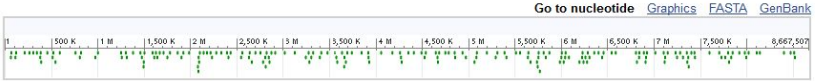
Genome Sequencing Projects

Chromosomes [1] Scaffolds or contigs [0] SRA or Traces [0] No data [1]

Organism	BioProject	Assembly	Status	Chrs	Plasmids	Size (Mb)	GC%	Gene	Protein
Streptomyces coelicolor	PRJNA13298	-	○	-	-	-	-	-	
Streptomyces coelicolor A3(2)	PRJNA57801, PRJNA242	ASM20383v1	●	1	2	9.05	72	8,300	8,153

Genome Region

Go to nucleotide Graphics FASTA GenBank



Other BioProjects

Epigenomics	3
Other	3
Transcriptome or Gene expression	30
Variation	1

Publications

1. Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in Streptomyces coelicolor. O'Rourke S, et al. Mol Microbiol 2009 Feb
2. Comparative genomics of Streptomyces avermitilis, Streptomyces cattleya, Streptomyces maritimus and Kitasatospora aureofaciens using a Streptomyces coelicolor microarray system. Hsiao NH, et al. Antonie Van Leeuwenhoek 2008 Jan-Feb
3. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining. Corre C, et al. Proc Natl Acad Sci U S A 2008 Nov 11

More...

Related information

BioProject
Gene
Protein Clusters
Components
Protein
PubMed
Taxonomy

Search details

```
"Streptomyces coelicolor"
[Organism]
```

Search | See more...

Recent activity

Turn Off Clear

- Streptomyces coelicolor — Genome
- streptomyces coelicolor (1) — Genome
- Homo sapiens — Genome
- Homo sapiens (1) — Genome
- SCO6094 (4) — Protein

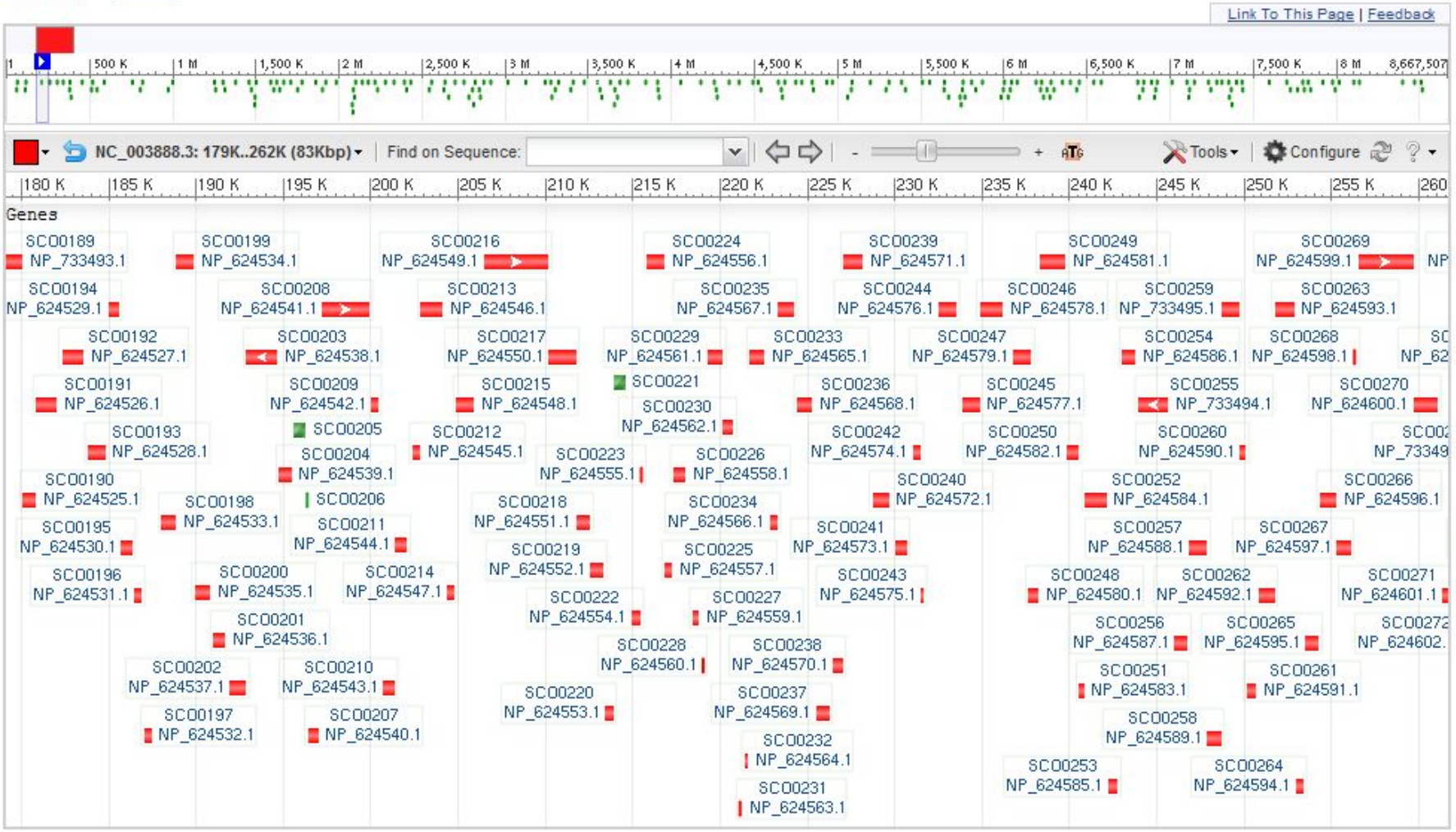
See more...

Genome

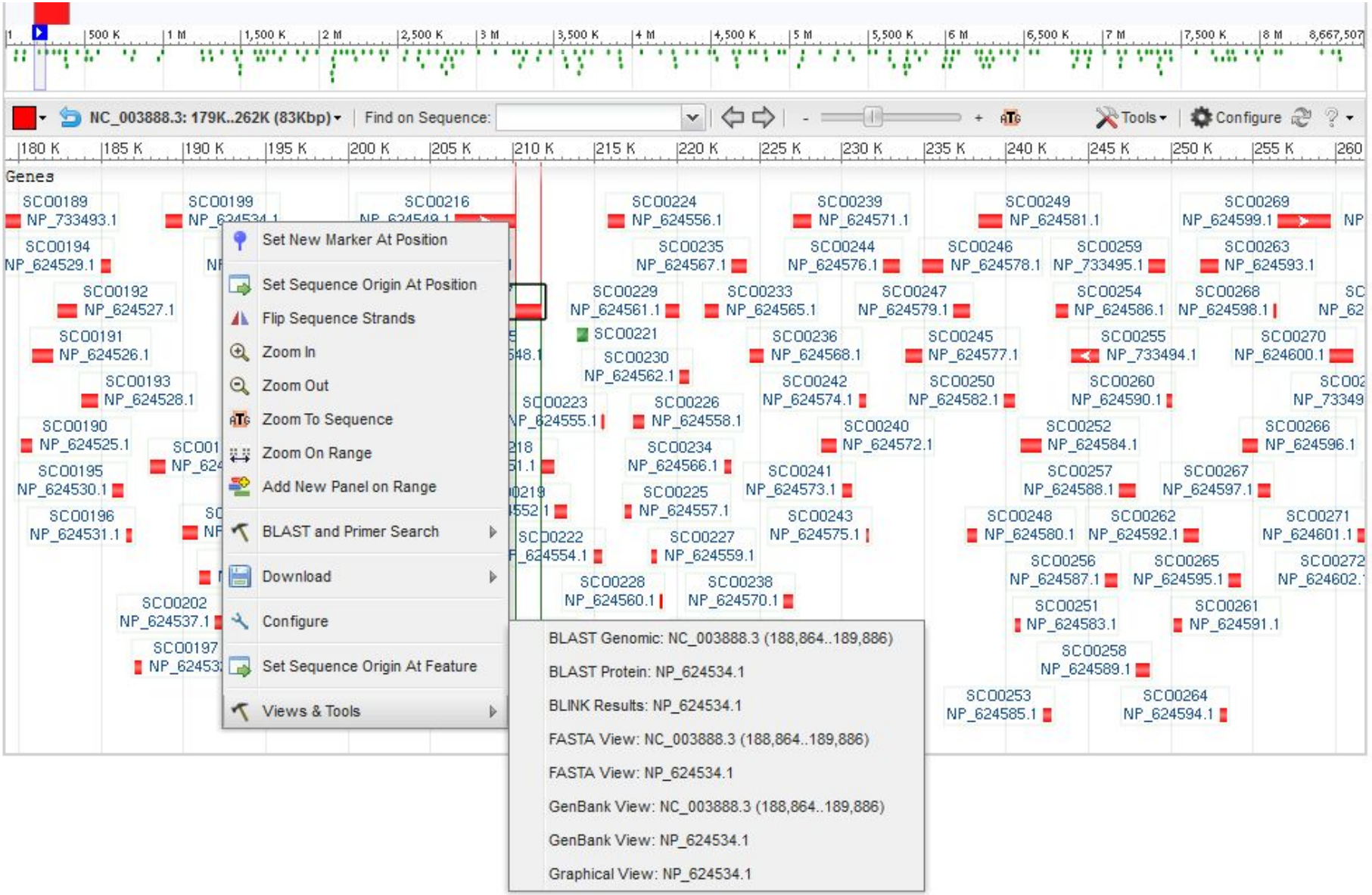
Streptomyces coelicolor A3(2) chromosome, complete genome

NCBI Reference Sequence: NC_003888.3

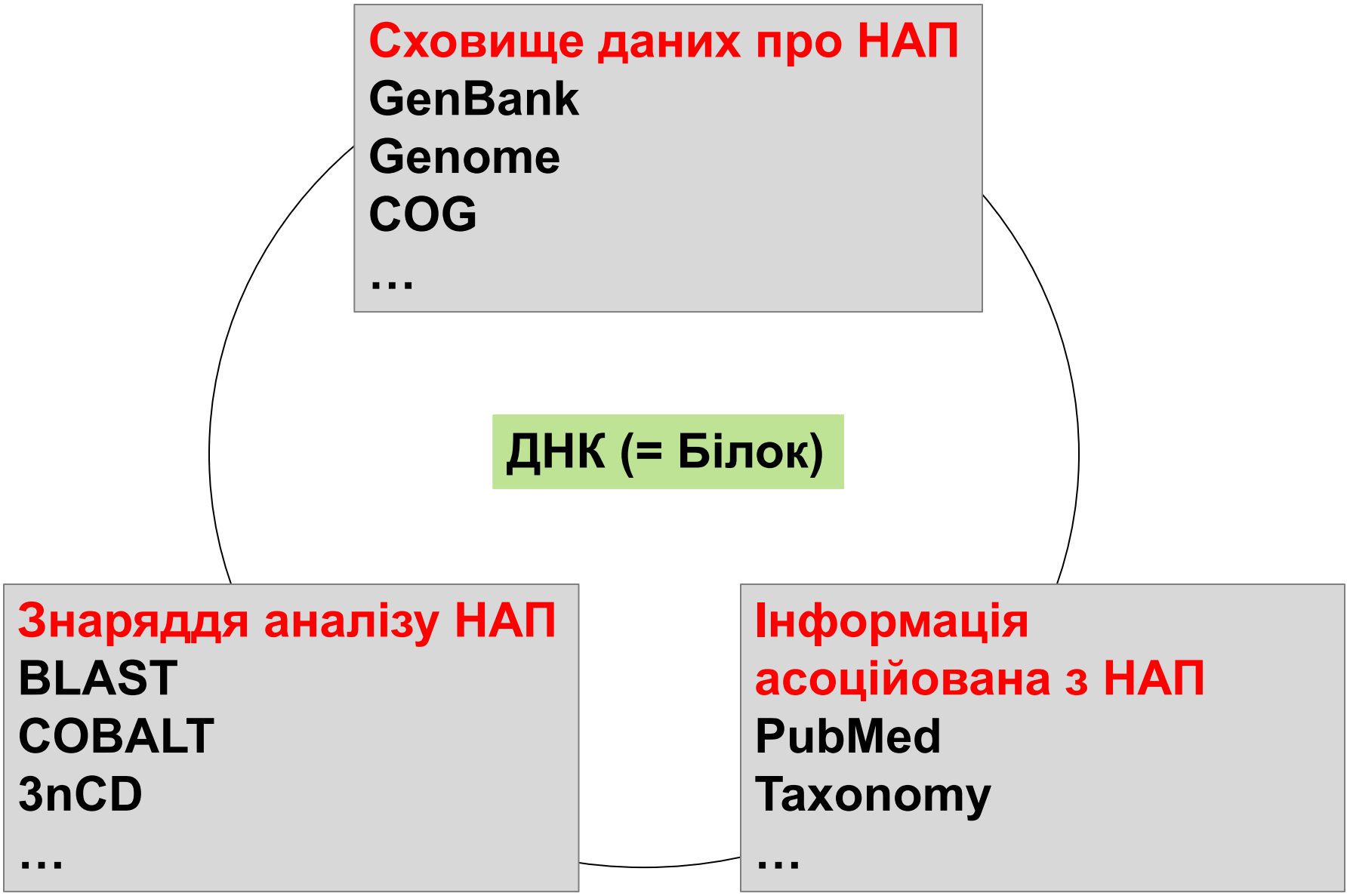
GenBank FASTA



Genome


NC_003888.3: 179K..262K (83Kbp)
Find on Sequence:
Tools
Configure
Genes
Set New Marker At Position
Set Sequence Origin At Position
Flip Sequence Strands
Zoom In
Zoom Out
Zoom To Sequence
Zoom On Range
Add New Panel on Range
BLAST and Primer Search
Download
Configure
Set Sequence Origin At Feature
Views & Tools
BLAST Genomic: NC_003888.3 (188,864..189,886)
BLAST Protein: NP_624534.1
BLINK Results: NP_624534.1
FASTA View: NC_003888.3 (188,864..189,886)
FASTA View: NP_624534.1
GenBank View: NC_003888.3 (188,864..189,886)
GenBank View: NP_624534.1
Graphical View: NP_624534.1

Модель NCBI


Сховище даних про НАП
GenBank
Genome
COG
…
ДНК (= Білок)
Знаряддя аналізу НАП
BLAST
COBALT
3nCD
…
Інформація асоційована з НАП
PubMed
Taxonomy
…