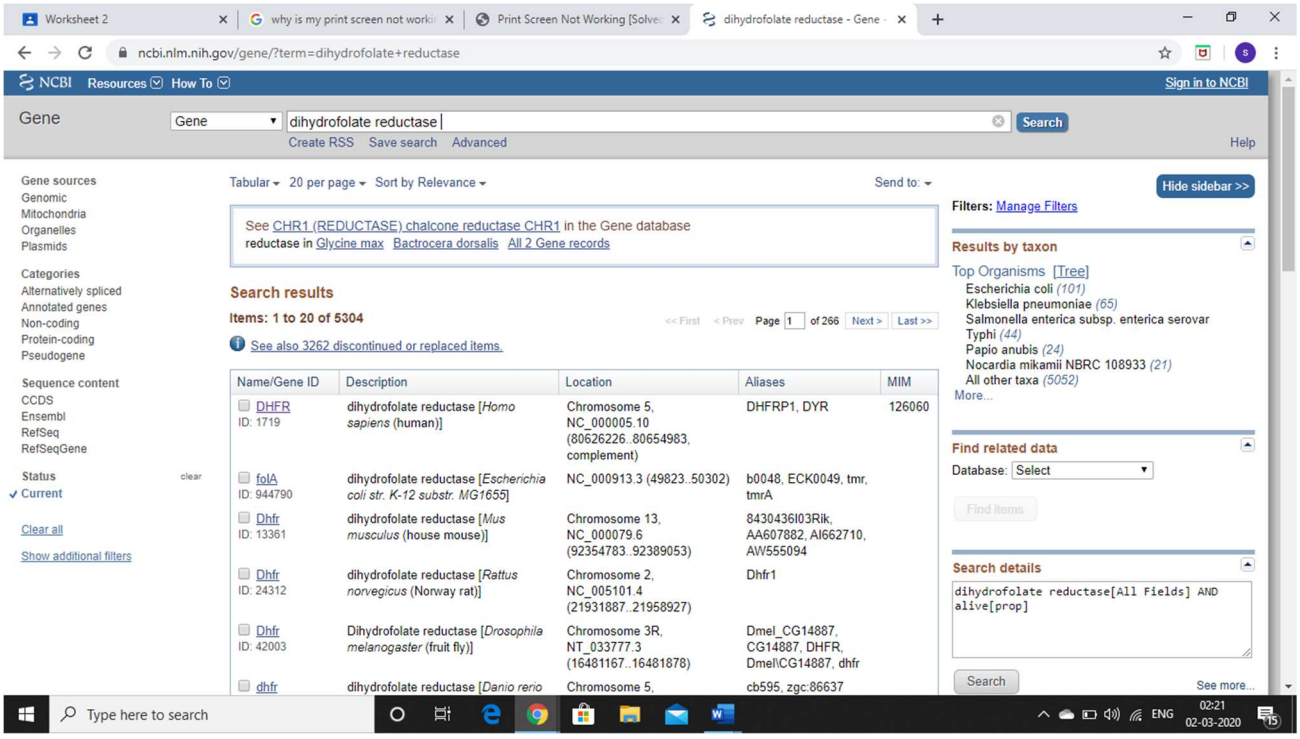
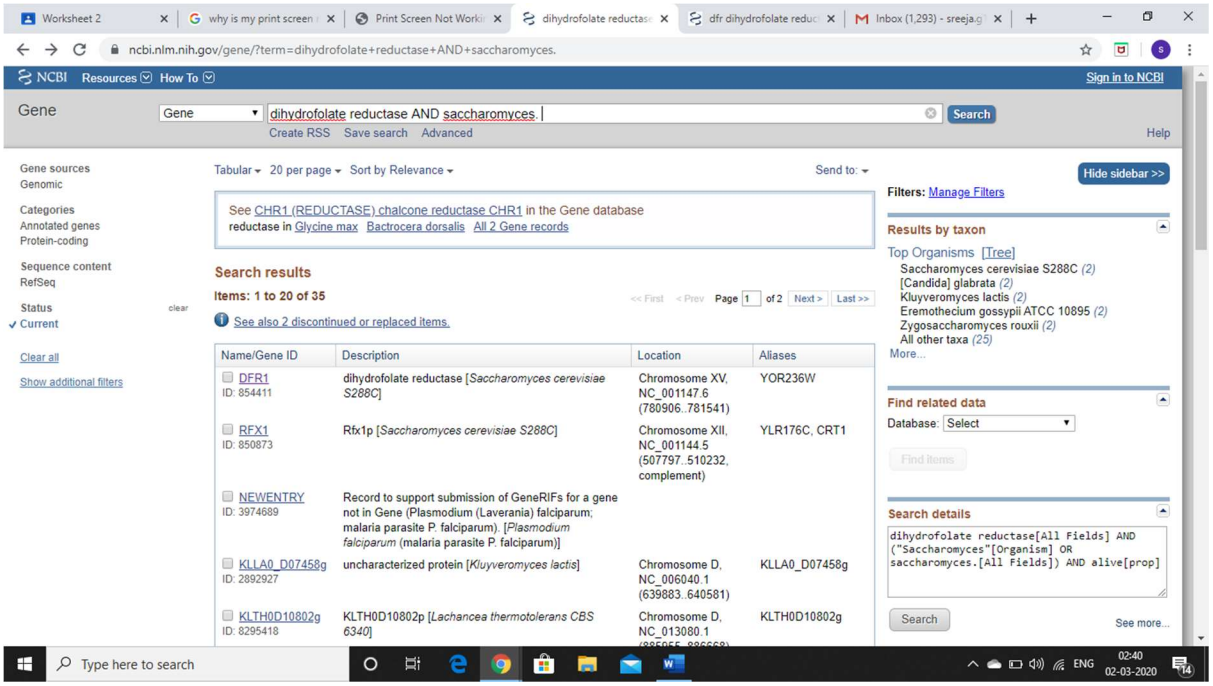
Bio-informatics assignment-2

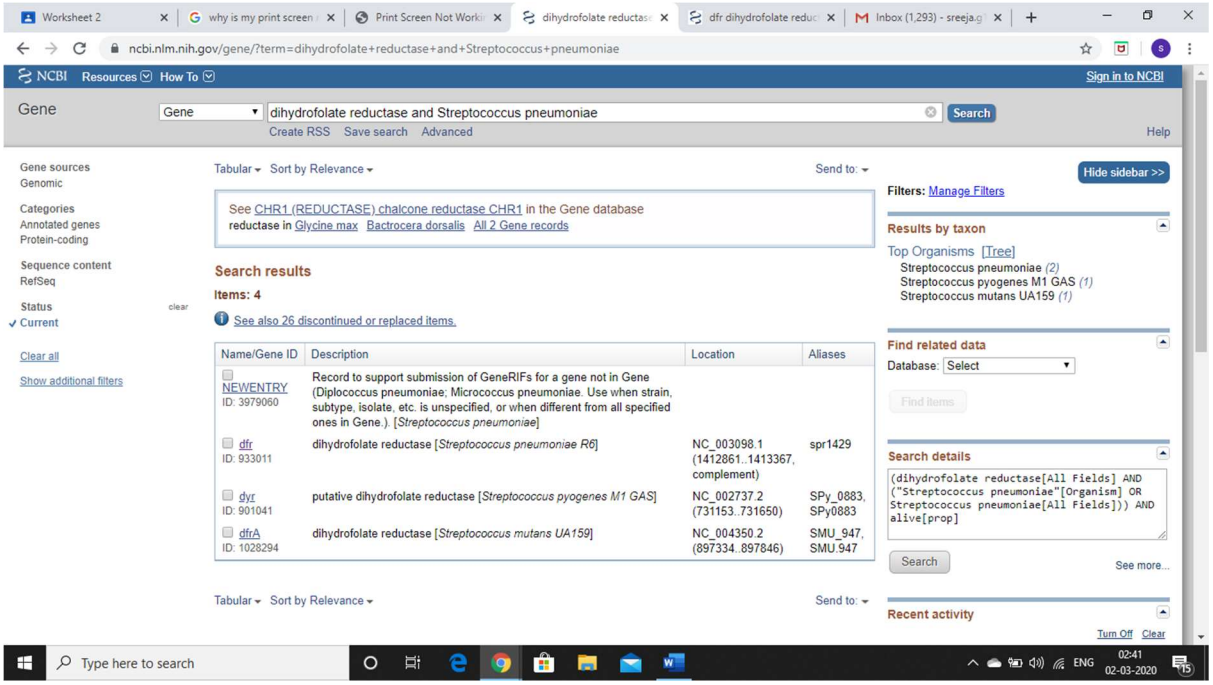
Sreeja Gaddamidi

S20170010047

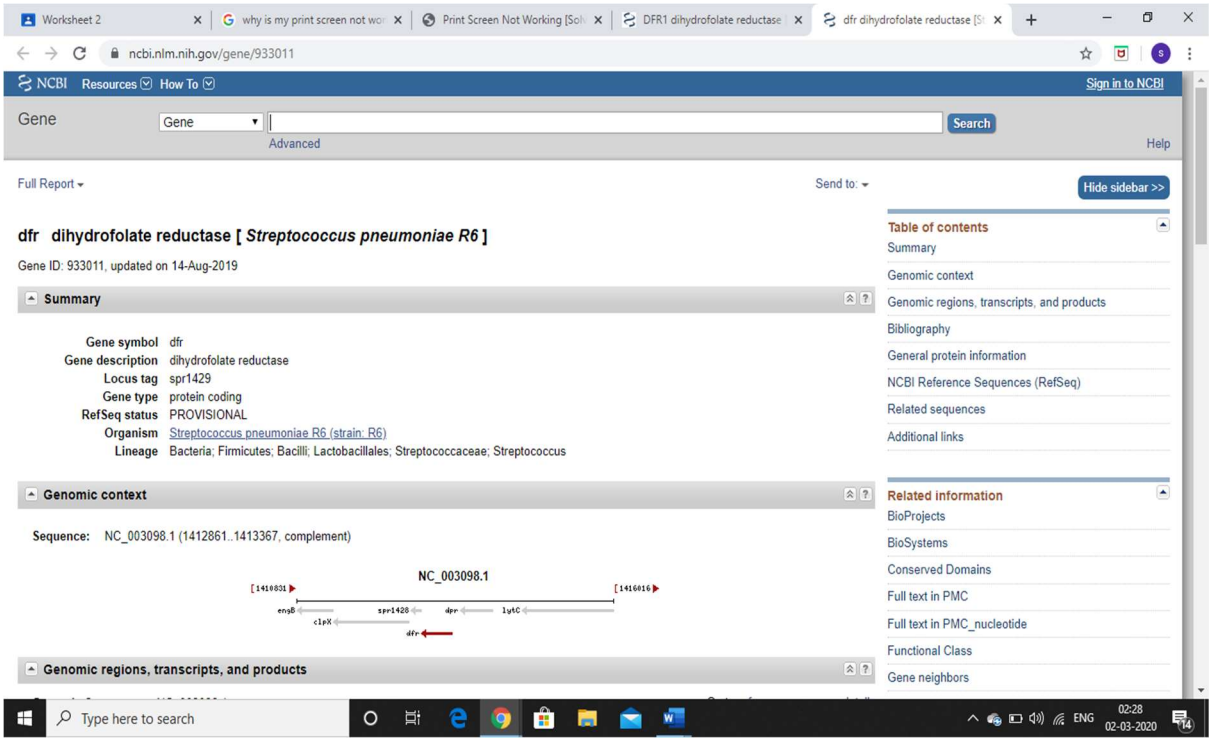
Task 1:

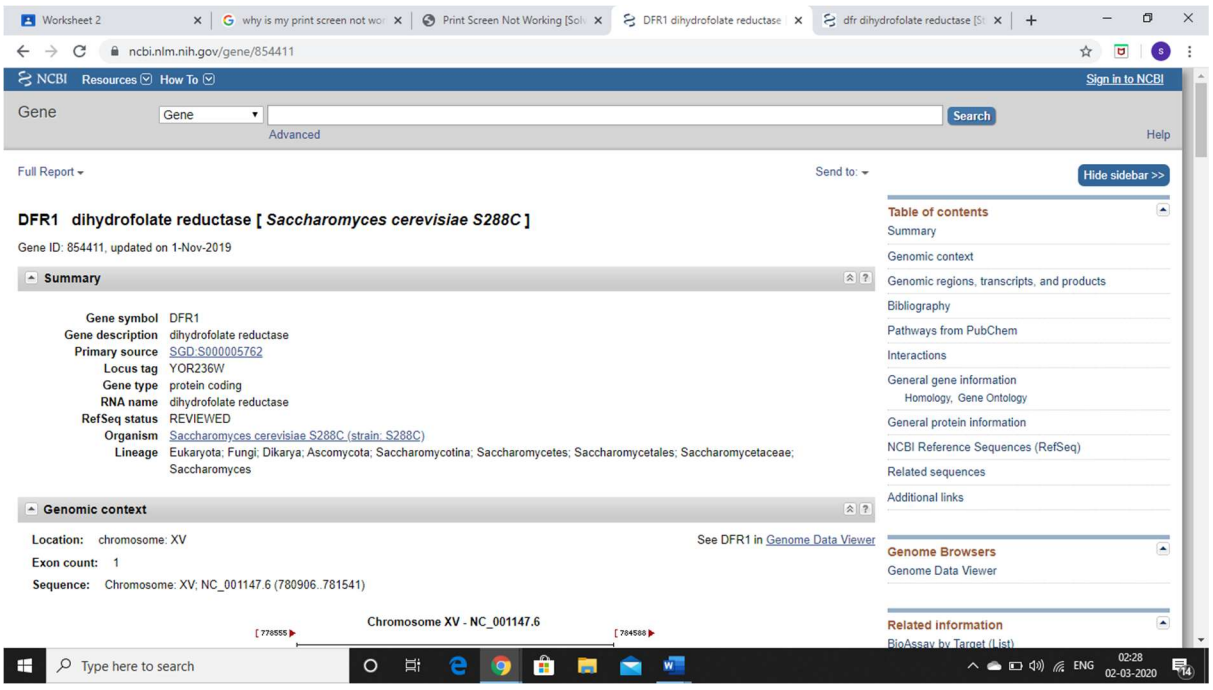



Worksheet 2
why is my print screen
Print Screen Not Worki
dihydrofolate reductase
dfr dihydrofolate reduc
Inbox (1,293) - sreeja.g
ncbi.nlm.nih.gov/gene/?term=dihydrofolate+reductase+AND+saccharomyces.
NCBI Resources How To
Sign in to NCBI
Gene
Gene
dihydrofolate reductase AND saccharomyces.
Search
Create RSS Save search Advanced
Help
Gene sources
Genomic
Categories
Annotated genes
Protein-coding
Sequence content
RefSeq
Status
Current
Clear all
Show additional filters
Tabular 20 per page Sort by Relevance
Send to:
Hide sidebar >>
Filters: Manage Filters
See CHR1 (REDUCTASE) chalcone reductase CHR1 in the Gene database
reductase in Glycine max Bactrocera dorsalis All 2 Gene records
Search results
Items: 1 to 20 of 35
<< First < Prev Page 1 of 2 Next > Last >>
See also 2 discontinued or replaced items.
Name/Gene ID
Description
Location
Aliases
DFR1
ID: 854411
dihydrofolate reductase [Saccharomyces cerevisiae S288C]
Chromosome XV, NC_001147.6 (780906..781541)
YOR236W
RFX1
ID: 850873
Rfx1p [Saccharomyces cerevisiae S288C]
Chromosome XII, NC_001144.5 (507797..510232, complement)
YLR176C, CRT1
NEWENTRY
ID: 3974689
Record to support submission of GeneRIFs for a gene not in Gene (Plasmodium (Laverania) falciparum; malaria parasite P. falciparum). [Plasmodium falciparum (malaria parasite P. falciparum)]
KLLA0_D07458g
ID: 2892927
uncharacterized protein [Kluyveromyces lactis]
Chromosome D, NC_006040.1 (639883..640581)
KLLA0_D07458g
KLTH0D10802g
ID: 8295418
KLTH0D10802p [Lachancea thermotolerans CBS 6340]
Chromosome D, NC_013080.1
KLTH0D10802g
Results by taxon
Top Organisms [Tree]
Saccharomyces cerevisiae S288C (2)
[Candida] glabrata (2)
Kluyveromyces lactis (2)
Eremothecium gossypii ATCC 10895 (2)
Zygosaccharomyces rouxii (2)
All other taxa (25)
More...
Find related data
Database: Select
Find items
Search details
dihydrofolate reductase[All Fields] AND ("Saccharomyces"[Organism] OR saccharomyces.[All Fields]) AND alive[prop]
Search
See more...
Type here to search
ENG
02:40
02-03-2020


Worksheet 2
why is my print screen
Print Screen Not Worki
dihydrofolate reductase
dfr dihydrofolate reduc
Inbox (1,293) - sreeja.g
ncbi.nlm.nih.gov/gene/?term=dihydrofolate+reductase+and+Streptococcus+pneumoniae
NCBI Resources How To
Sign in to NCBI
Gene
Gene
dihydrofolate reductase and Streptococcus pneumoniae
Search
Create RSS Save search Advanced
Help
Gene sources
Genomic
Categories
Annotated genes
Protein-coding
Sequence content
RefSeq
Status
Current
Clear all
Show additional filters
Tabular Sort by Relevance
Send to:
Hide sidebar >>
Filters: Manage Filters
See CHR1 (REDUCTASE) chalcone reductase CHR1 in the Gene database
reductase in Glycine max Bactrocera dorsalis All 2 Gene records
Search results
Items: 4
See also 26 discontinued or replaced items.
Name/Gene ID
Description
Location
Aliases
NEWENTRY
ID: 3979060
Record to support submission of GeneRIFs for a gene not in Gene (Diplococcus pneumoniae; Micrococcus pneumoniae. Use when strain, subtype, isolate, etc. is unspecified, or when different from all specified ones in Gene.). [Streptococcus pneumoniae]
dfr
ID: 933011
dihydrofolate reductase [Streptococcus pneumoniae R6]
NC_003098.1 (1412861..1413367, complement)
spr1429
dyr
ID: 901041
putative dihydrofolate reductase [Streptococcus pyogenes M1 GAS]
NC_002737.2 (731153..731650)
SPy_0883, SPy0883
dfrA
ID: 1028294
dihydrofolate reductase [Streptococcus mutans UA159]
NC_004350.2 (897334..897846)
SMU_947, SMU.947
Tabular Sort by Relevance
Send to:
Results by taxon
Top Organisms [Tree]
Streptococcus pneumoniae (2)
Streptococcus pyogenes M1 GAS (1)
Streptococcus mutans UA159 (1)
Find related data
Database: Select
Find items
Search details
(dihydrofolate reductase[All Fields] AND ("Streptococcus pneumoniae"[Organism] OR Streptococcus pneumoniae[All Fields])) AND alive[prop]
Search
See more...
Recent activity
Turn Off Clear
Type here to search
ENG
02:41
02-03-2020

Task 2:





Task 3:

The genomic context of Streptococcus pneumonia R6:

Gene symbol - DFR

Organism (Scientific name)- Streptococcus pneumonia R6 (strain: R6)

Gene ID – 933011

Gene Type- Protein coding

Lineages- Bacteria, Bacilli, etc

Genomic Sequence Code – NC_003098.1

Red color shows the directionality which is from left to right for DFR.

The genomic context of Saccharomyces cerevisiaeS2886 are:

Gene symbol - DFR

Organism (Scientific name)- Streptococcus pneumonia R6 (strain: R6)

Gene ID – 933011

Gene Type- Protein coding

Lineages- Bacteria, Bacilli, etc

Genomic Sequence Code – NC_003098.1

Red color shows the directionality which is from left to right for DFR

Task 4:

Saccharomyces cerevisiae (Yeast) has 211 amino acids.

Streptococcus Pneumoniae has 168 amino acids.

Task 5:

Two main features of Saccharomyces cerevisiae :

```
FEATURES             Location/Qualifiers
     source          1..636
                     /organism="Saccharomyces cerevisiae S288C"
                     /mol_type="genomic DNA"
                     /strain="S288C"
                     /db_xref="taxon:559292"
                     /chromosome="XV"
     gene            <1..>636
                     /gene="DFR1"
                     /locus_tag="YOR236W"
                     /db_xref="GeneID:854411"
     mRNA            <1..>636
                     /gene="DFR1"
                     /locus_tag="YOR236W"
                     /product="dihydrofolate reductase"
                     /transcript_id="NM_001183655.1"
                     /db_xref="GeneID:854411"
```

Two main features of Streptococcus Pneumoniae:

```
FEATURES             Location/Qualifiers
     source          1..168
                     /organism="Streptococcus pneumoniae R6"
                     /strain="R6"
                     /db_xref="taxon:171101"
     Protein         1..168
                     /product="dihydrofolate reductase"
                     /EC_number="1.5.1.3"
                     /calculated_mol_wt=19630
     Region          4..165
                     /region_name="DHFR_1"
                     /note="Dihydrofolate reductase; pfam00186"
                     /db_xref="CDD:333908"
```

Task 6:

a. S288C

1. In the locus info for protein we can again see the protein code as well as it we come to know that this protein is made up of 211 amino acids.
2. 8 References for work related to this protein is given in the database. Some references include 'The nucleotide sequence for of Saccharomyces cerevisiae chromosome XV'(PUBMED: 9169874) from the journal Nature and 'Life with 6000 genes '(PUBMED: 8849441) from the Science Journal.
3. Although, the protein as a whole is a dhfr, it is actually amino acids from 8 to 208 which is actually called as the 'DHFR' region. The second figure below highlights the 'DHFR' region.
4. S288C protein starts with 'm' amino acid and ends with 'k' amino acid. DHFR region starts with 'i' amino acid and ends with 'y' amino acid

b) R6 .

1. R6 protein is made up of 168 amino acids which is lower than S288C.
2. 3 references for work related to this protein is provided in the database. References for these are official government research based references according to the R6 protein details. This is under the NCBI Genome Project which aims at collecting genomic information from various organisms and studying them in order to collect data.

GenPept

dihydrofolate reductase [Saccharomyces cerevisiae S288C]

NCBI Reference Sequence: NP_014879.1

Identical Proteins FASTA Graphics

Go to:

```
LOCUS       NP_014879               211 aa            linear   PLN 01-NOV-2019
DEFINITION  dihydrofolate reductase [Saccharomyces cerevisiae S288C].
ACCESSION   NP_014879
VERSION     NP_014879.1
DBLINK      BioProject: PRJNA128
DBSOURCE    REFSEQ: accession NM_001183655.1
KEYWORDS    RefSeq.
SOURCE      Saccharomyces cerevisiae S288C
  ORGANISM  Saccharomyces cerevisiae S288C
            Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;
            Saccharomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (residues 1 to 211)
  AUTHORS   Dujon,B., Albermann,K., Aldea,M., Alexandraki,D., Ansorge,W.,
            Arino,J., Benes,V., Bohn,C., Bolotin-Fukuhara,M., Bordonne,R.,
            Boyer,J., Camasses,A., Casamayor,A., Casas,C., Cheret,G.,
            Cziepluch,C., Daignan-Fornier,B., Dang,D.V., de Haan,M., Delius,H.,
            Durand,P., Fairhead,C., Feldmann,H., Gaillon,L., Kleine,K. et al.
  TITLE     The nucleotide sequence of Saccharomyces cerevisiae chromosome XV
  JOURNAL   Nature 387 (6632 SUPPL), 98-102 (1997)
   PUBMED   9169874
REFERENCE   2  (residues 1 to 211)
  AUTHORS   Goffeau,A., Barrell,B.G., Bussey,H., Davis,R.W., Dujon,B.,
            Feldmann,H., Galibert,F., Hoheisel,J.D., Jacq,C., Johnston,M.,
            Louis,E.J., Mewes,H.W., Murakami,Y., Philippsen,P., Tettelin,H. and
            Oliver,S.G.
  TITLE     Life with 6000 genes
  JOURNAL   Science 274 (5287), 546 (1996)
   PUBMED   8849441
REFERENCE   3  (residues 1 to 211)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2019) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   4  (residues 1 to 211)
  CONSRTM   Saccharomyces Genome Database
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2015) Department of Genetics, Stanford
            University, Stanford, CA 94305-5120, USA
  REMARK    Protein update by submitter
REFERENCE   5  (residues 1 to 211)
```

Streptococcus pneumoniae R6 chromosome, complete genome

NCBI Reference Sequence: NC_003098.1

FASTA Graphics

Go to:

```
LOCUS       NC_003098               507 bp    DNA     linear   CON 11-JAN-2017
DEFINITION  Streptococcus pneumoniae R6 chromosome, complete genome.
ACCESSION   NC_003098 REGION: 1412861..1413367
VERSION     NC_003098.1
DBLINK      BioProject: PRJNA57859
            Assembly: GCF_000007045.1
KEYWORDS    RefSeq.
SOURCE      Streptococcus pneumoniae R6
  ORGANISM  Streptococcus pneumoniae R6
            Bacteria; Firmicutes; Bacilli; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.Richard., Rosteck,P.R. Jr.,
            Skatrud,P.L. and Glass,J.I.
  TITLE     Genome of the bacterium Streptococcus pneumoniae strain R6
  JOURNAL   J. Bacteriol. 183 (19), 5709-5717 (2001)
   PUBMED   11544234
REFERENCE   2  (bases 1 to 507)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2001) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (bases 1 to 507)
  AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.Richard., Rosteck,P.R. Jr.,
            Skatrud,P.L. and Glass,J.I.
```

Task 7:

FASTA for Streptococcus Pneumoniae Protein:

>>NP_359022.1 dihydrofolate reductase [Streptococcus pneumoniae R6]

MTKKIVAIWAQDEEGLIGKENRLPWHLPAELQHFKETTLNHAILMGRVTFDGMGRRLLPKRETLILTRNP

EEKIDGVATFQDVQSVLDWYQDQEKNLYIIGGKQIFQAFEPYLDEVIVTHIHARVEGDTYFPEELDLSLF

ETVSSKFYAKDEKNPYDFTIQYRKRKEV

FASTA for Saccharomyces Protein:

>NP_014879.1 dihydrofolate reductase [Saccharomyces cerevisiae S288C]

MAGGKIPIVGIVACLQPEMGIGFRGGLPWRLPSEMKYFRQVTSLTKDPNKKNALIMGRKTWESIPPKFRP

LPNRMNVIISRSFKDDFVHDKERSIVQSNSLANAIMNLESNFKEHLERIYVIGGGEVYSQIFSITDHWLI

TKINPLDKNATPAMDTFLDAKKLEEVFSEQDPAQLKEFLPPKVELPETDCDQRYSLEEKGYCFEFTLYNR

K

FASTA for Streptococcus Pneumoniae Nucleotide:

>NC_003098.1:1412861-1413367 Streptococcus pneumoniae R6 chromosome, complete genome

TTAGACTTCCTTTCTCTTGCGGTATTGGATGGTAAAATCATAAGGATTCTTCTCATCTTTGGCGTAAAAT
TTGCTTGAAACTGTCTCAAAAAGAGACAAGTCAAGCTCTTCAGGGAAATAGGTATCTCCTTCCACCCGAG
CATGAATGTGAGTGACAATCACTTCATCAAGGTAAGGTTCAAAAGCCTGAAAAATTTGCTTCCCACCGAT
AATGTAGAGATTCTTTTCTTGATCCTGATACCAGTCAAGAACAGACTGGACGTCCTGAAAAGTAGCAACC
CCATCTATCTTTTCTTCCGGATTACGCGTCAAAATCAGGGTTTCCCGTTTTGGAAGCAAGCGACGCCCCA
TCCCATCAAAGGTCACACGCCCCATCAAGATAGCATGATTCAGAGTTGTTTCTTTAAAGTGCTGCAATTC
TGCTGGCAAATGCCAAGGCAGACGATTTTCCTTACCAATCAAACCCTCTTCATCCTGGGCCCAAATAGCT
ACGATTTTCTTAGTCAT

FASTA for Saccharomyces Nucleotide:

>NC_001147.6:780906-781541 Saccharomyces cerevisiae S288C chromosome XV, complete sequence

ATGGCTGGAGGAAAGATTCCTATTGTAGGAATTGTGGCATGTTTACAGCCGGAGATGGGGATAGGATTTC
GTGGAGGTCTACCATGGAGGTTGCCCAGTGAAATGAAGTATTTCAGACAGGTCACTTCATTGACGAAAGA
TCCAAACAAAAAAAAATGCTTTGATAATGGGAAGGAAGACATGGGAATCCATACCGCCCAAGTTTCGCCCA
CTGCCCAATAGAATGAATGTCATTATATCGAGAAGCTTCAAGGACGATTTTGTCCACGATAAAGAGAGAT
CAATAGTCCAAAGTAATTCATTGGCAAACGCAATAATGAACCTAGAAAGCAATTTTAAGGAGCATCTGGA
AAGAATCTACGTGATTGGGGGTGGCGAAGTTTATAGTCAAATCTTCTCCATTACAGATCATTGGCTCATC
ACGAAAATAAATCCATTAGATAAAAACGCAACTCCTGCAATGGACACTTTCCTTGATGCGAAGAAATTGG
AAGAAGTATTTAGCGAGCAAGATCCGGCCCAGCTGAAAGAATTTCTTCCCCCTAAAGTAGAGTTGCCCGA
AACAGACTGTGATCAACGCTACTCGCTGGAAGAAAAAGGTTATTGCTTCGAATTCACTCTATACAATCGT
AAATGA