

Assignment - 2 Bio-Informatics

Name : Neelakanta Sriram
Roll.No : S20170010102

Task - 1:

Exact Number of Entries

The screenshot shows the NCBI Gene search interface. The search term 'dihydrofolate reductase' has been entered into the search bar. The results page displays 20 entries per page, sorted by relevance. A summary box at the top indicates 'See CHR1 (REDUCTASE) chalcone reductase CHR1 in the Gene database' and lists 'reductase 2 in Glycine max', 'Bactrocera dorsalis', and 'All 2 Gene records'. The main results table includes columns for Name/Gene ID, Description, Location, Aliases, and MIM. The table lists various genes from different species, including Homo sapiens, Escherichia coli, Mus musculus, Rattus norvegicus, Drosophila melanogaster, Danio rerio, and Galus gallus. The total number of results is 5304, with 266 pages available.

Name/Gene ID	Description	Location	Aliases	MIM
DHFR	dihydrofolate reductase [Homo sapiens (human)] ID: 1719	Chromosome 5, NC_000005.10 (80626226..80654983, complement)	DHFRP1, DYR	126060
tflA	dihydrofolate reductase [Escherichia coli str. K-12 substr. MG1655]	NC_000913.3 (498073..50302)	b0040, ECK0040, Imt, ImA	
dhfr	dihydrofolate reductase [Mus musculus (house mouse)] ID: 944790	Chromosome 13, NC_000079.6 (9254783..9238905)	84304360309k, AA070782, AI662710, AW555094	
dhfr	dihydrofolate reductase [Rattus norvegicus (Norway rat)] ID: 13361	Chromosome 2, NC_005101.4 (21933887..21998627)	DHFR	
dhfr	dihydrofolate reductase [Drosophila melanogaster (fruit fly)] ID: 42003	Chromosome 3R, NT_033777.3 (16481167..16481878)	Dmel_CG14887, CG14887, DHFR, DmelCG14887, dhfr	
dhfr	dihydrofolate reductase [Danio rerio (zebrafish)] ID: 81882	Chromosome 5, NC_007116.7 (26138013..26145917)	c595, zgc:86637	
DHFR2	dihydrofolate reductase 2 [Homo sapiens (human)] ID: 200895	Chromosome 3, NC_000003.12 (84057922..84063389, complement)	DHFR1, DHFR4	616588
DFRL1	dihydrofolate reductase [Saccharomyces cerevisiae S288C] ID: 854411	Chromosome XV, NC_001147.6 (789096..781541)	YOR236W	
DHFRP1	dihydrofolate reductase pseudogene 1 [Homo sapiens (human)] ID: 573971	Chromosome 18, NC_000018.10 (26167862..26171357, complement)		
dhfr	dihydrofolate reductase [Escherichia coli] ID: 13906554	NC_019082.1 (4613..5110)	D616_p72001	
DHFR	dihydrofolate reductase [Galus gallus (chicken)] ID: 427317	Chromosome Z, NC_006127.5 (64257309..64273299)		
AT2G21550	Bifunctional dihydrofolate reductase/thymidylate synthase [Arabidopsis thaliana (thale cress)] ID: 816694	Chromosome 2, NC_003071.7 (9226917..9230205)	AT2G21550, F2G1.18, F2G1_18	

Task - 2:

A. Dihydrofolate reductase AND Saccharomyces.

NCBI Resources How To

Gene Gene dihydrofolate reductase AND saccharomyces

Create RSS Save search Advanced

Send to: Search Hide sidebar >>

Search results

Items: 1 to 20 of 37 Selected: 1

Name/Gene ID	Description	Location	Aliases
<input checked="" type="checkbox"/> DFR1 ID: 654411	dihydrofolate reductase [Saccharomyces cerevisiae S288C]	Chromosome XV, NC_001147.6 (789096..781541)	YOR236W
<input type="checkbox"/> REF1 ID: 650673	Rh1p [Saccharomyces cerevisiae S288C]	Chromosome XII, NC_001144.5 (567797..510232, complement)	YLR176C, CRT1
<input type="checkbox"/> NEWENTRY ID: 3974689	Record to support submission of GeneRIFs for a gene not in Gene (Plasmodium (Laveranii) falciparum; malaria parasite P. falciparum). [Plasmodium falciparum] (malaria parasite P. falciparum)]	Chromosome D, NC_006040.1 (639883..640581)	KLLA0_D07458g
<input type="checkbox"/> KLLA0_D07458g ID: 2895297	uncharacterized protein [Kluyveromyces lactis]	Chromosome D, NC_013060.1 (889595..889668)	KLT00D10802g
<input type="checkbox"/> KLT00D10802g ID: 8295418	KLT00D10802p [Lachancea thermotolerans CBS 6340]	Chromosome C, NC_006069.1 (177351..1774100, complement)	YALI0_C12771g
<input type="checkbox"/> YALI0_C12771g ID: 2905913	YALI0C12771p [Yarrowia lipolytica CLIB122]	Chromosome C, NC_006045.2 (1448332..1448952, complement)	DEHA2C16456g
<input type="checkbox"/> DEHA2C16456g ID: 2906017	DEHA2C16456p [Debaryomyces hansenii CBS767]	LALAO_501e03978g	
<input type="checkbox"/> LALAO_501e03378g ID: 34683488	uncharacterized protein [Lachancea lanzensteinii]	Chromosome F, NC_012995.1	ZYR00F06930g
<input type="checkbox"/> ZYR00F06930g	hypothetical protein [Zygosaccharomyces rouxii]		

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 (27) See more...

Find related data Database: Select Find items

Search details dihydrofolate reductase[All Fields] AND ("Saccharomyces" [Organism] OR saccharomyces[All Fields])

Search See more...

B . Dihydrofolate reductase AND Streptococcus pneumoniae

NCBI Resources How To

Gene Gene Streptococcus pneumoniae AND dihydrofolate reductase

Create RSS Save search Advanced

Send to: Search Hide sidebar >>

Search results

Items: 4 Selected: 1

See CHR1 (REDUCTASE) chloramphenicol reductase CHR1 in the Gene database
reductase in Glycine max Bactrocera dorsalis All 2 Gene records

Send to: Filters: Manage Filters

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 ("Streptococcus pneumoniae"[Organism] OR Streptococcus pneumoniae[All Fields]) AND dihydrofolate reductase[All Fields] AND alive[prop]

Search See more...

Name/Gene ID	Description	Location	Aliases
<input type="checkbox"/> NEWENTRY ID: 3979080	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]	NC_003098.1 (1412861..1413367, complement)	spr1429
<input checked="" type="checkbox"/> dfr ID: 933011	dihydrofolate reductase [Streptococcus pneumoniae R9]	NC_002737.2 (731153..731650)	SPY_0883, SPY0883
<input type="checkbox"/> dfr ID: 901041	putative dihydrofolate reductase [Streptococcus pyogenes M1 GAS]	NC_004305.2 (897334..897846)	SMU_947, SMU947
<input type="checkbox"/> dfr ID: 1028294	dihydrofolate reductase [Streptococcus mutans UA159]		

Tabular Sort by Relevance

Task - 3

Information Gathered From Bars :

----> The genomic context of **Streptococcus pneumoniae R6**:

Gene symbol - DFR

Scientific name - Streptococcus pneumoniae 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 cerevisiae S288C** are:

Gene symbol - DFR

Scientific name - Streptococcus pneumoniae R6 (strain: R6)

Gene ID - 933011

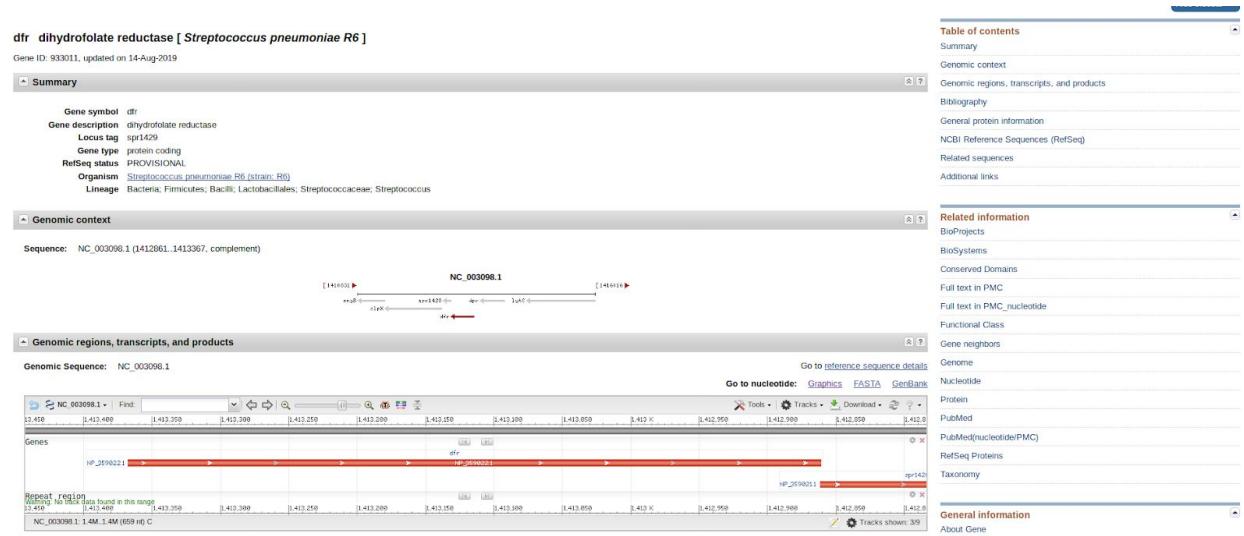
Gene Type - Protein coding

Lineages - Bacteria, Bacilli, etc

Genomic Sequence Code - NC_003098.1

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

Streptococcus :



Saccharomyces :

DFR1 dihydrofolate reductase [*Saccharomyces cerevisiae* S288C]

Gene ID: 854411, updated on 1-Nov-2019

Summary

Gene symbol DFR1
 Gene description dihydrofolate reductase
 Primary gene ID 854411
 Locus tag YOR238W
 Gene type protein coding
 RNA name dihydrofolate reductase
 RefSeq status REVIEWED
 Organism *Saccharomyces cerevisiae* S288C (strain: S288C)
 Lineage Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces

Genomic context

Location: chromosome: XV
 Exon count: 1
 Sequence: Chromosome: XV; NC_001147.6 (780906..781541)

See DFR1 in [Genome Data Viewer](#)

Genomic regions, transcripts, and products

Genomic Sequence: NC_001147.6

Chromosome XV - NC_001147.6

Go to reference sequence details | Go to nucleotide: Graphics FASTA GenBank

Genes, RefSeq propagation from SGD, annotation version R64-2-1

Go to nucleotide: Graphics FASTA GenBank

Bibliography

Related articles in PubMed

1. Nucleotide sequence of the dihydrofolate reductase gene of *Saccharomyces cerevisiae*.

Summary
Genomic context
Genomic regions, transcripts, and products
Bibliography
Pathways from PubChem
Interactions
General gene information
 Homology, Gene Ontology
General protein information
NCBI Reference Sequences (RefSeq)
Related sequences
Additional links
Genome Browsers
Genome Data Viewer
Related information
 BioAssay by Target (List)
 BioAssay by Target (Summary)
 BioProjects
BioSystems
 Conserved Domains
 Full text in PMC_nucleotide
 Functional Class
 Gene neighbors
 Genome
 GEO Profiles
 HomoloGene
Nucleotide
Probe
Protein
PubMed

Task - 4

Saccharomyces (Yeast) has **211** amino acids.
Streptococcus (Bacteria) has **168** amino acids.

Task - 5

Features of *Saccharomyces* (Yeast) are :

FEATURES	source	Location/Qualifiers		Gene
		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"		
		/products="dihydrofolate reductase"		
		/transcript_id="NM_001183655.1"		
		/db_xref="GeneID:854411"		
CDS		1..636		
		/gene="DFR1"		
		/locus_tag="YOR236W"		
		/EC_number="1.5.1.3"		
		/experiment="EXISTENCE:curator inference:GO:0005829		
		cytosol [PMID:1427091 PMID:2827121 PMID:2838385]		
		/experiment="EXISTENCE:direct assay:GO:0003729 mRNA binding [PMID:21124907]"		
		/experiment="EXISTENCE:direct assay:GO:0004146 dihydrofolate reductase activity [PMID:2838385]"		
		/experiment="EXISTENCE:direct assay:GO:0005737 cytoplasm [PMID:14562095]"		
		/experiment="EXISTENCE:direct assay:GO:0005739 mitochondrion [PMID:14562095]"		
		/experiment="EXISTENCE:mutant phenotype:GO:0004146 dihydrofolate reductase activity [PMID:1427091 PMID:2827121]"		
		/experiment="EXISTENCE:mutant phenotype:GO:0046452 dihydrofolate metabolic process [PMID:1427091]"		
		/experiment="EXISTENCE:mutant phenotype:GO:0046654 tetrahydrofolate biosynthetic process [PMID:1427091]"		
		/note="Dihydrofolate reductase involved in tetrahydrofolate biosynthesis; required for respiratory metabolism; mutation is functionally complemented by human DHFR"		
		/codon_start=1		
		/products="dihydrofolate reductase"		
		/protein_id="NP_014879.1"		
		/db_xref="GeneID:854411"		
		/db_xref="SGD:S000005762"		
		/translation="MAGGKIPIVGVACLOPEMIGFRGLPWRLPSEMKYFRQVTSL TKDPNKKNALIMGRKTWE1PPKFRPLPNRMNVIIISRSFKDDFVHDKERSIVQSNSLA NAIMNLESNFKEHLERIYVGGEVYSQIFSITDHWLITKINPLDKNATPAMDTEFLDA KKLEEVFSEQDPAQLKEFLPPKVELPETDCDQRYSLEEKGYCFEFTLYNRK"		
ORIGIN		1 atggctggag gaaagattcc tatttgtttaga atttgtggcat gtttacagcc ggagatgggg		

Features of *Streptococcus* (Bacteria) are:

Task - 6

Saccharomyces : S288C

This protein is made up of 211 amino acids. Some references related to proteins in the database include '**The nucleotide sequence for of *Saccharomyces cerevisiae* chromosome XV**' from the journal Nature and '**Life with 6000 genes**' from the Science Journal.

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. 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

Saccharomyces cerevisiae S288C chromosome XV, complete sequence

NCBI Reference Sequence: NC_001147.6

[FASTA](#) [Graphics](#)

Go to:

LOCUS NC_001147 636 bp DNA linear CON 01-NOV-2019
DEFINITION Saccharomyces cerevisiae S288C chromosome XV, complete sequence.
ACCESSION NC_001147 REGION: 780906..781541
VERSION NC_001147.6
DBLINK BioProject: PRJNA128
Assembly: GCF_000146045.2
KEYWORDS RefSeq.
SOURCE Saccharomyces cerevisiae S288C
ORGANISM Saccharomyces cerevisiae S288C
Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;
Saccharomycetes; Saccharomycetales; Saccharomycetaceae;
Saccharomyces.
REFERENCE 1 (bases 1 to 636)
AUTHORS Dujon,B., Albermann,K., Aldea,M., Alexandraki,D., Ansorge,W.,
Arino,J., Benes,V., Bohn,C., Bolotin-Fukuhara,M., Bordone,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 (bases 1 to 636)
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., Philippson,P., Tettelin,H. and
Oliver,S.G.
TITLE Life with 6000 genes
JOURNAL Science 274 (5287), 546 (1996)
PUBMED 8849441
REFERENCE 3 (bases 1 to 636)
CONSRM NCBI Genome Project
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2019) National Center for Biotechnology
Information, NIH, Bethesda, MD 20894, USA
REFERENCE 4 (bases 1 to 636)
CONSRM 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 (bases 1 to 636)
CONSRM Saccharomyces Genome Database
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2012) Department of Genetics, Stanford

Streptococcus:

R6 .

This is made up of 168 amino acids which is lower than S288C.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.

⚠ Due to the large size of this record, sequence and annotated features are not shown. Use the "Customize view" panel to change the display.

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., Blaszcak,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)
CONSRM 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., Blaszcak,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.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
COMMENT REVIEWED [REFSEQ](#): This record has been curated by NCBI staff. The
reference sequence was derived from [AF007317](#)

Task 7:

FASTA for Streptococcus Pneumoniae Protein:

```
>>NP_359022.1 dihydrofolate reductase [Streptococcus pneumoniae R6]  
MTKKIVAIWAQDEEGLIGKENRLPWHLPALQHFKETTLNHAILGRVTFDGMGRLLPKRETLI  
LTRNPEEKIDGVATFQDVQSVDWYQDQEKNLYIIGGKQIFQAFEPYLDEVIVTHIHARVEGDTY  
FPEELDLSLFETVSSKFYAKDEKNPYDFTIQYRKRKEV
```

FASTA for Saccharomyces Protein:

```
>NP_014879.1 dihydrofolate reductase [Saccharomyces cerevisiae S288C]  
MAGGKIPIVGIVACLQPEMGIGFRGLPWRLPSEMKYFRQVTSLKDPNKKNALIMGRKTWESI  
PPKFRPLPNRMNVIISRSFKDDFVHDKERSIVQSNSLANAIMNLESNFKEHLERIYVIGGEVYS  
QIFSITDHWLITKINPLDKNATPAMDTFLDAKKLEEVFSEQDPAQLKEFLPPKVELPETDCDQRY  
SLEEKGYCFEFTLYNRK
```

FASTA for Streptococcus Pneumoniae Nucleotide:

```
>NC_003098.1:1412861-1413367 Streptococcus pneumoniae R6  
chromosome, complete genome  
TTAGACTTCCTTCTCTTGCGGTATTGGATGGTAAAATCATAAGGATTCTTCATCTTGCG  
GTAAAATTGCTTGAAACTGTCTCAAAAAGAGACAAGTCAGCTCTCAGGGAAATAGGTAT  
CTCCTCCACCCGAGCATGAATGTGAGTGACAATCACTTCATCAAGGTAAGTTCAAAGC  
CTGAAAAATTGCTTCCCACCGATAATGTAGAGATTCTTCTTGATCCTGATACCAGTCA  
GAACAGACTGGACGTCTGAAAAGTAGCAACCCCCATCTATCTTCTCGGATTACCGCGT  
CAAAATCAGGGTTCCGTTGGAAAGCAAGCGACGCCCATCCCATCAAAGGTACACG  
CCCCATCAAGATAGCATGATTCAGAGTTCTTAAAGTGCTGCAATTCTGCTGGCAAAT  
GCCAAGGCAGACGATTTCTTACCAATCAAACCCTCTTCATCCTGGGCCAAATAGCTAC  
GATTTCCTTAGTCAT
```

FASTA for Saccharomyces Nucleotide:

```
>NC_001147.6:780906-781541 Saccharomyces cerevisiae S288C  
chromosome XV,complete sequence  
ATGGCTGGAGGAAAGATT CCTATTGTAGGAATTGTGGCATGTTACAGCCGGAGATGGGG  
ATAGGATTTCGTGGAGGTCTACCAGGAGGTTGCCAGTGAAATGAAGTATTTCAGACAGG  
TCACTTCATTGACGAAAGATCCAACAAAAAAATGCTTGATAATGGGAAGGAAGACATGG  
GAATCCATACCGGCCAAGTTCGCCCCACTGCCAATAGAATGAATGTCATTATATCGAGAA  
GCTTCAAGGACGATTTGTCCACGATAAAGAGAGATCAATAGTCAAAGTAATTCTGGCA  
AACGCAATAATGAACCTAGAAAGCAATTTTAAGGAGCATCTGGAAAGAATCTACGTGATTG  
GGGTGGCGAAGTTTAGTCAATCTTCTCCATTACAGATCATTGGCTCATCACGAAAATA
```

AATCCATTAGATAAAAACGCAACTCCTGCAATGGACACTTCCCTGATGCGAAGAAATTGGA
AGAAGTATTAGCGAGCAAGATCCGGCCCAGCTGAAAGAATTCTTCCCCCTAAAGTAGAG
TTGCCCGAAACAGACTGTGATCACGCTACTCGCTGGAAGAAAAGGTTATTGCTTCGAAT
TCACTCTATAACAATCGTAAATGA