



# An Introduction to NCBI's Bioinformatics Resources

Dr. Medha Devare

[mhd6@cornell.edu](mailto:mhd6@cornell.edu)

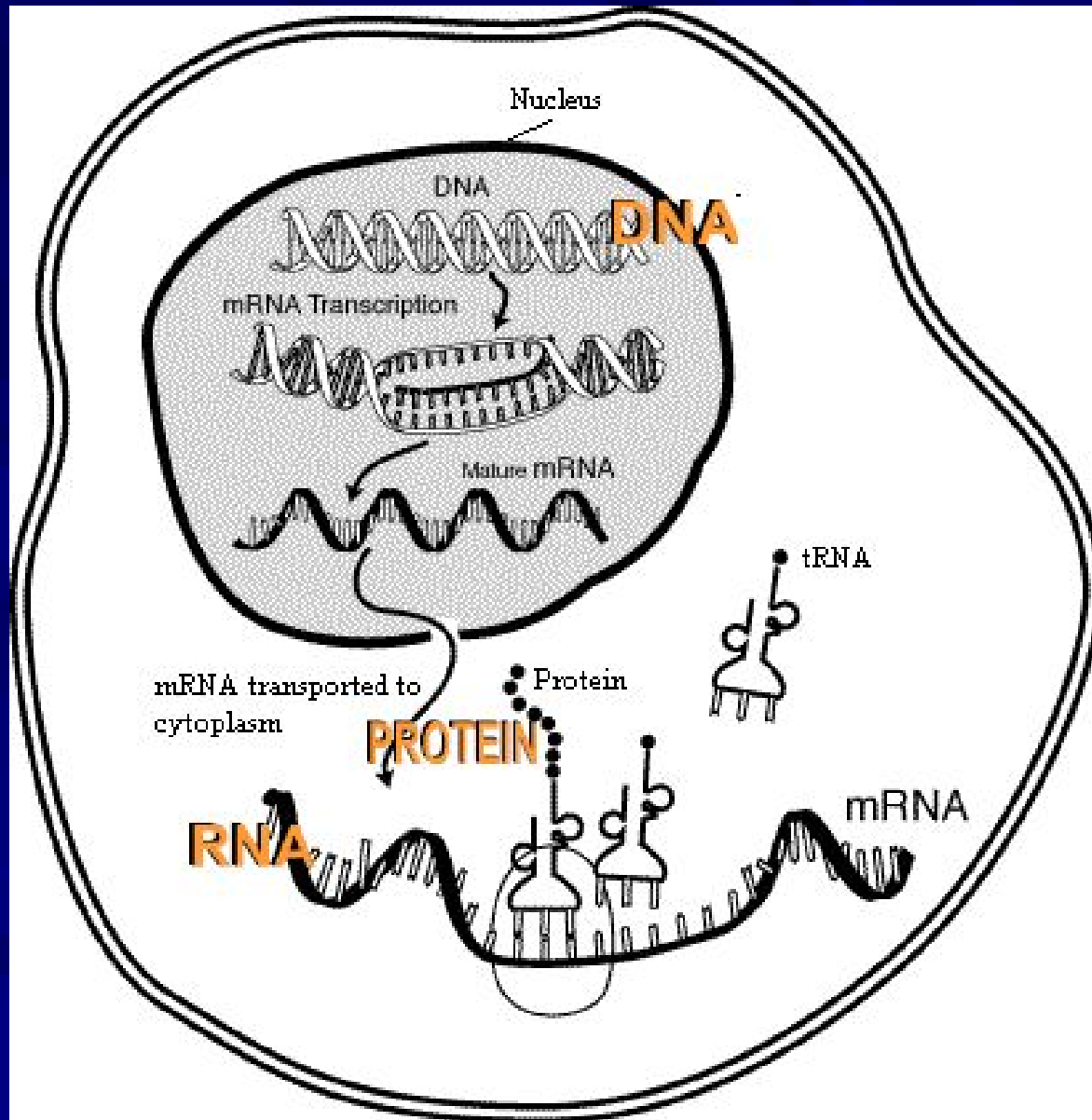
Life Sciences/Bioinformatics Specialist  
Albert R. Mann Library  
Cornell University, Ithaca, NY 14853

**USAIN 2006: *Delivering Information for the New Life Sciences***  
**October 7, 2006**

**Part I: Introduction to DNA Sequencing**

**Part II: Data Mining in Bioinformatics**

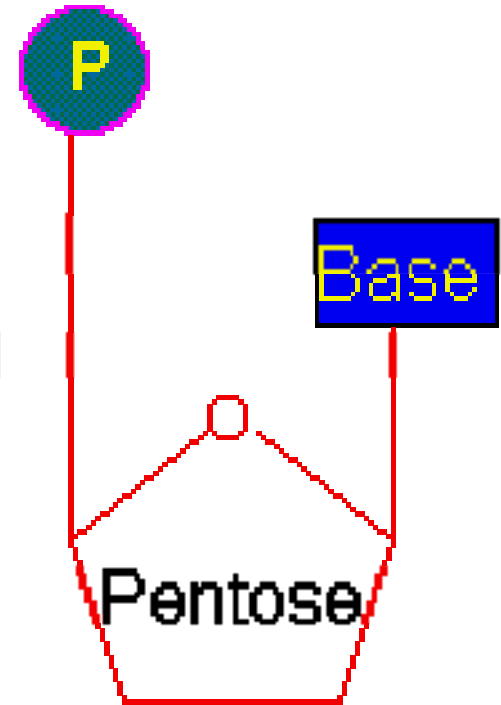
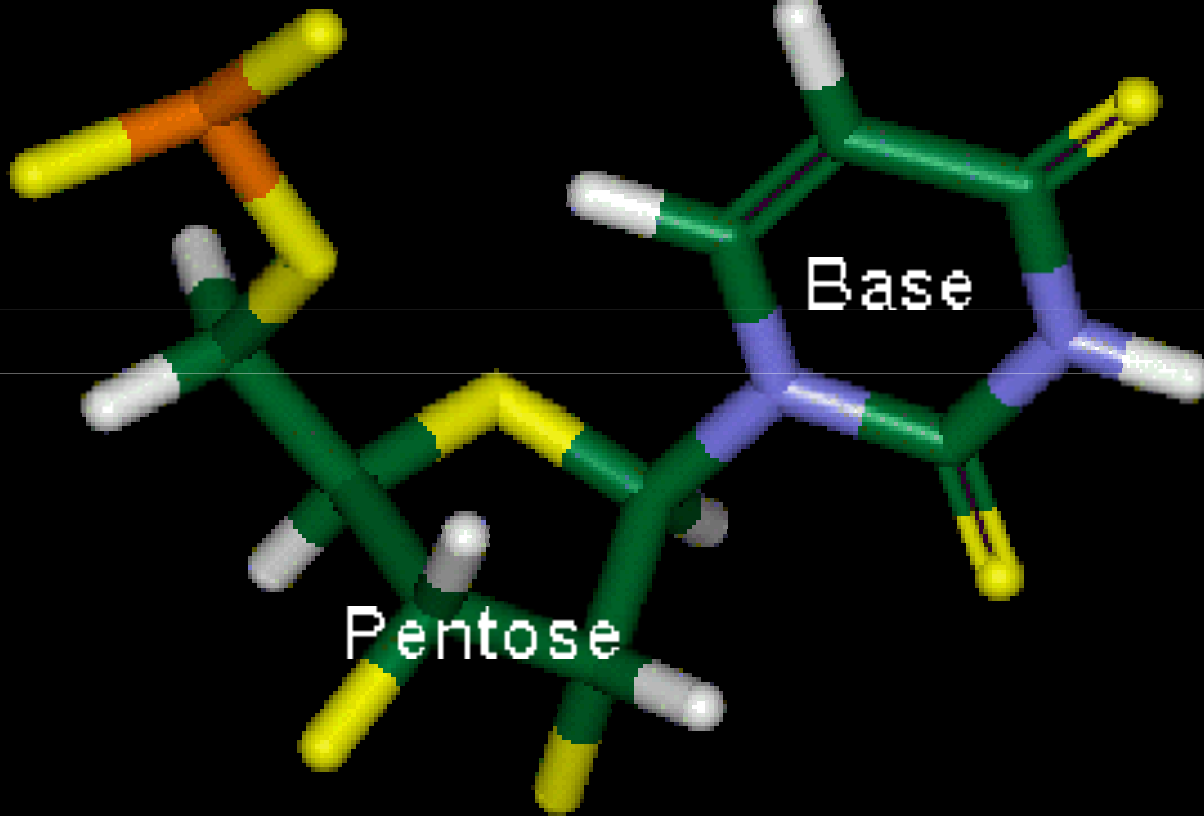
# CENTRAL DOGMA OF BIOLOGY



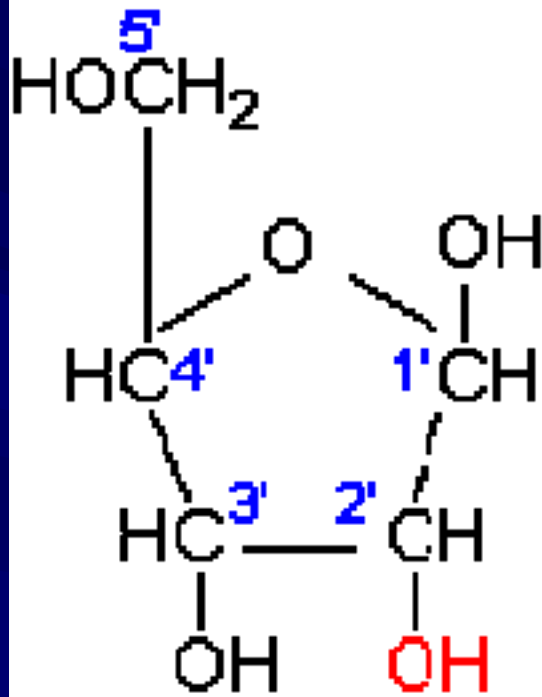
# NUCLEOTIDES

**Nucleotide = phosphate + pentose sugar + base**

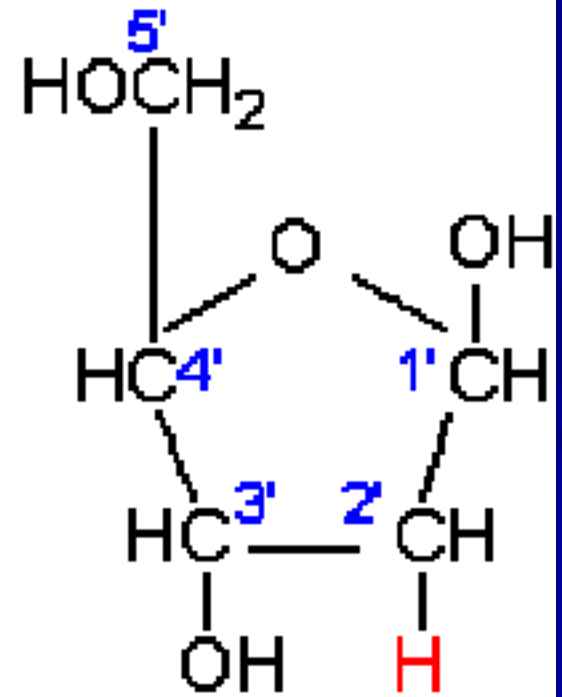
Phosphate



# PENTOSE SUGARS



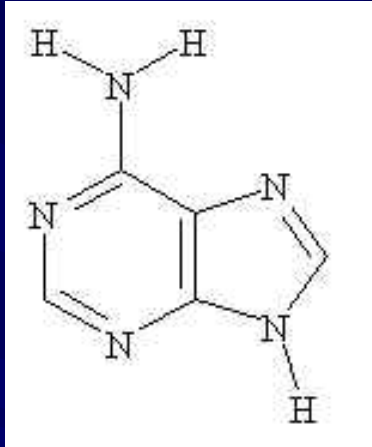
Ribose  
(in RNA)



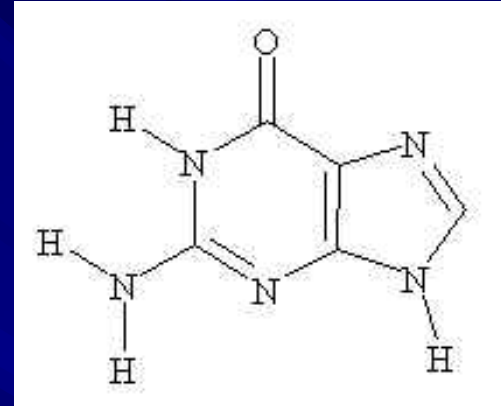
2'-Deoxyribose  
(in DNA)

# NITROGENOUS BASES

## Purines

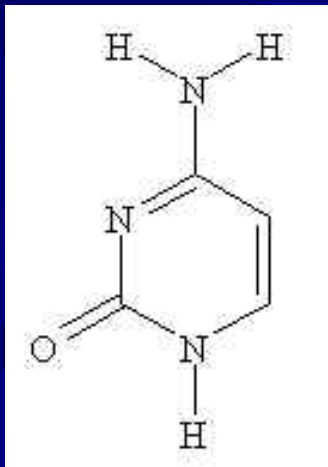


Adenine

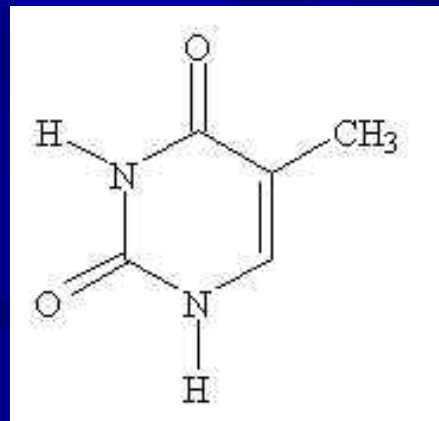


Guanine

## Pyrimidines



Cytosine

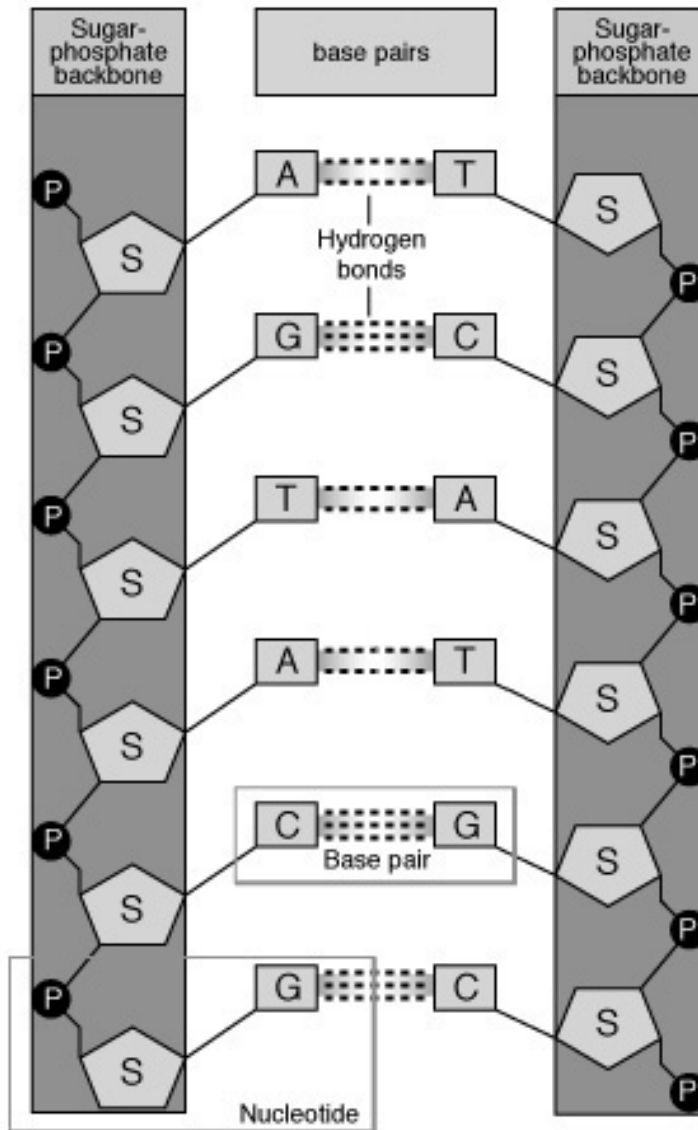


Thymine



Uracil (RNA only)

# STRUCTURE OF DNA

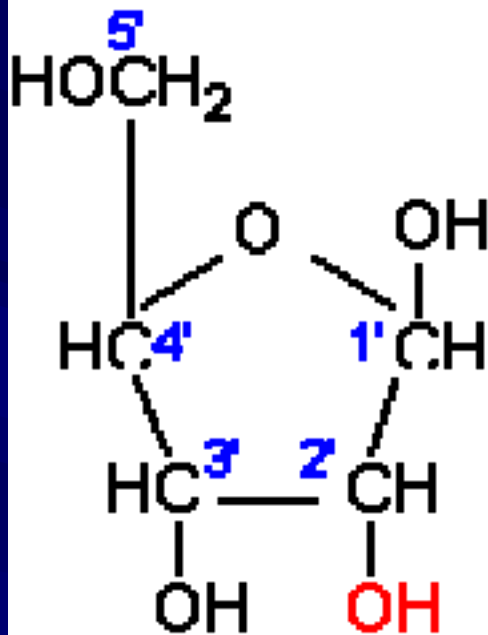


# DNA REPLICATION

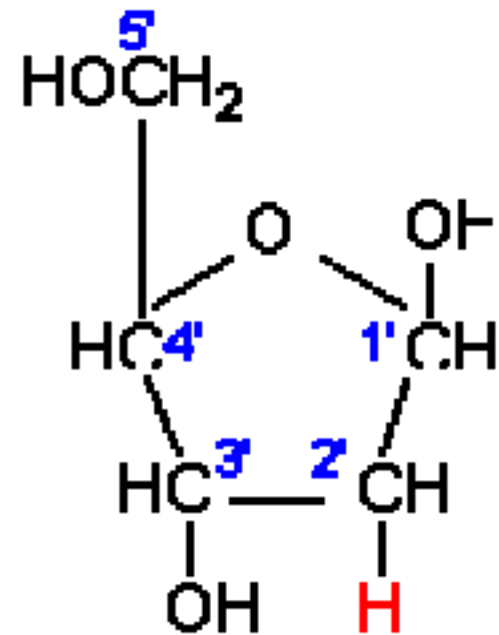
<http://www.ncc.gmu.edu/dna/repanim.htm>



# DNA SEQUENCING



Ribose  
(in RNA)



2'-Deoxyribose  
(in DNA)

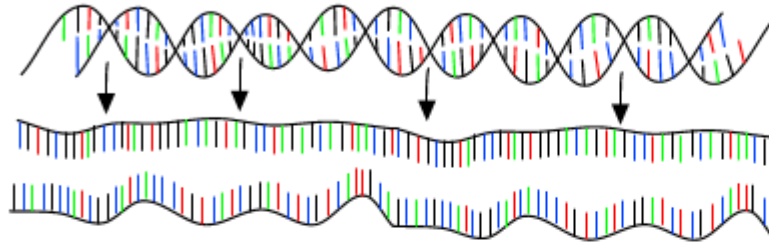
# DNA SEQUENCING

## Sequencing

30 cycles of 3 steps :

### Step 1 : denaturation

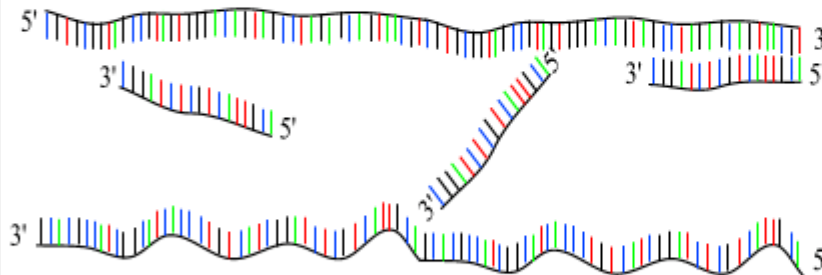
1 minut 94 °C



### Step 2 : annealing

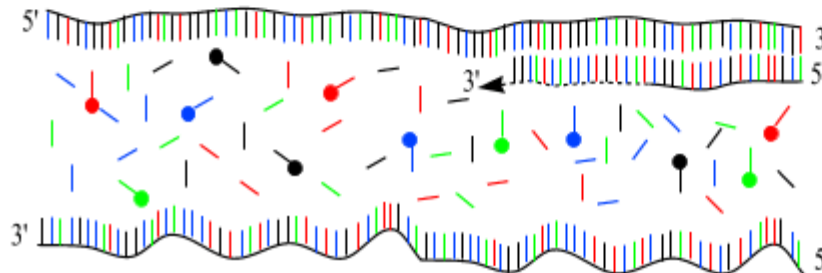
15 seconds 50 °C

1 primer !!!!



### Step 3 : extension

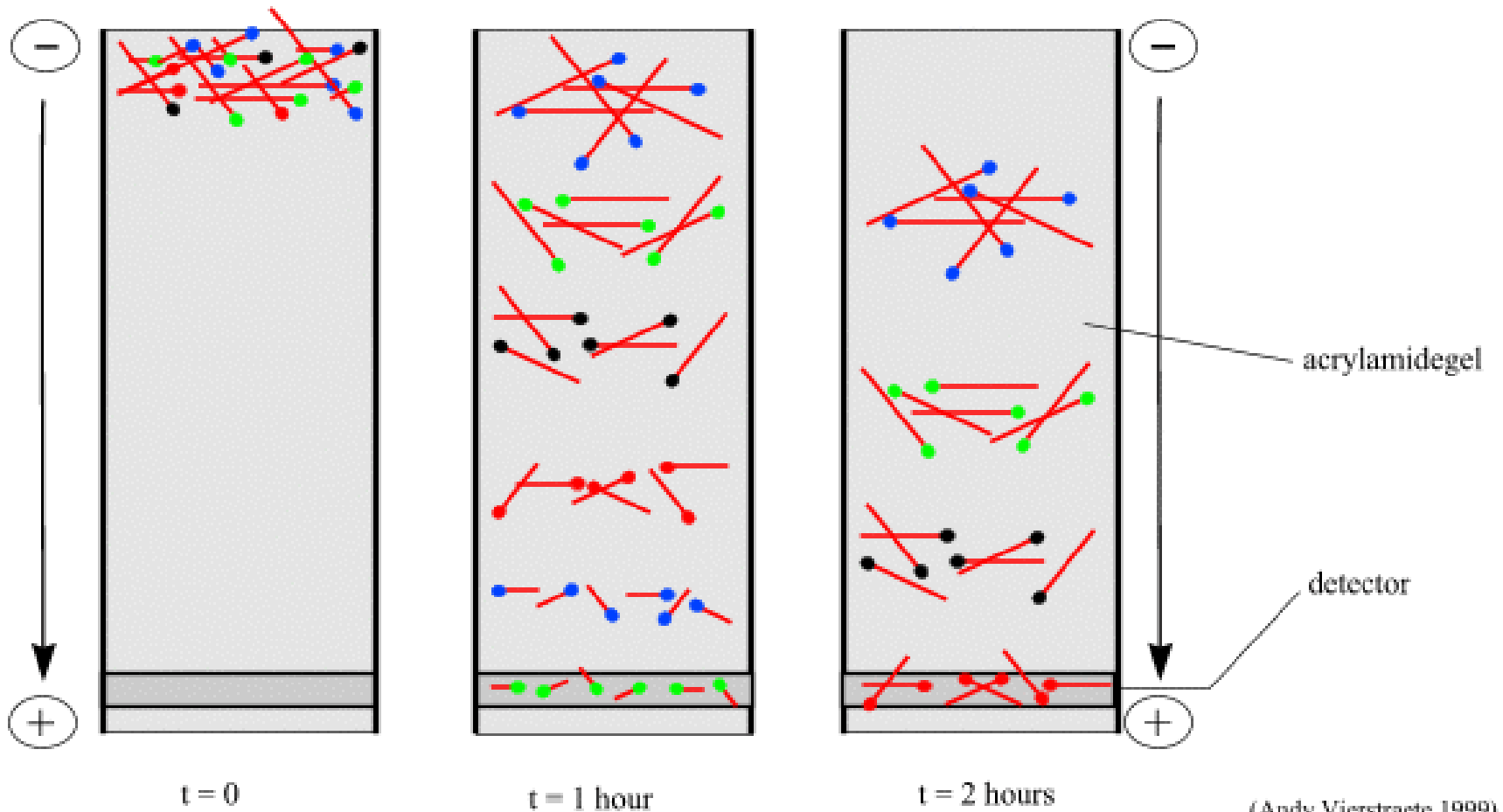
4 minutes 60 °C  
mixture of dNTP's |  
and ddNTP's ↓



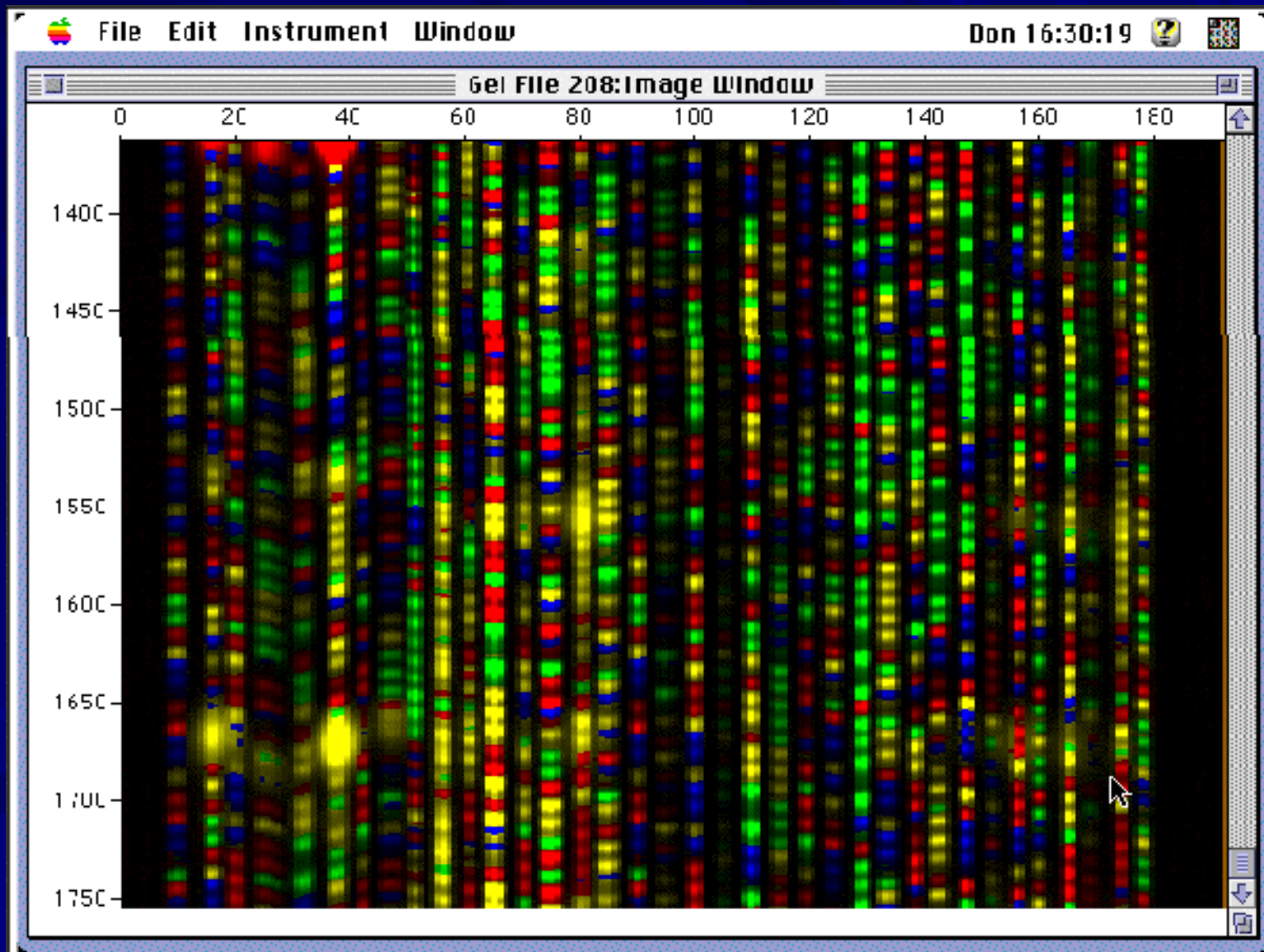
(Andy Vierstraete 1999)

# DNA SEQUENCING

## Gel electrophoresis



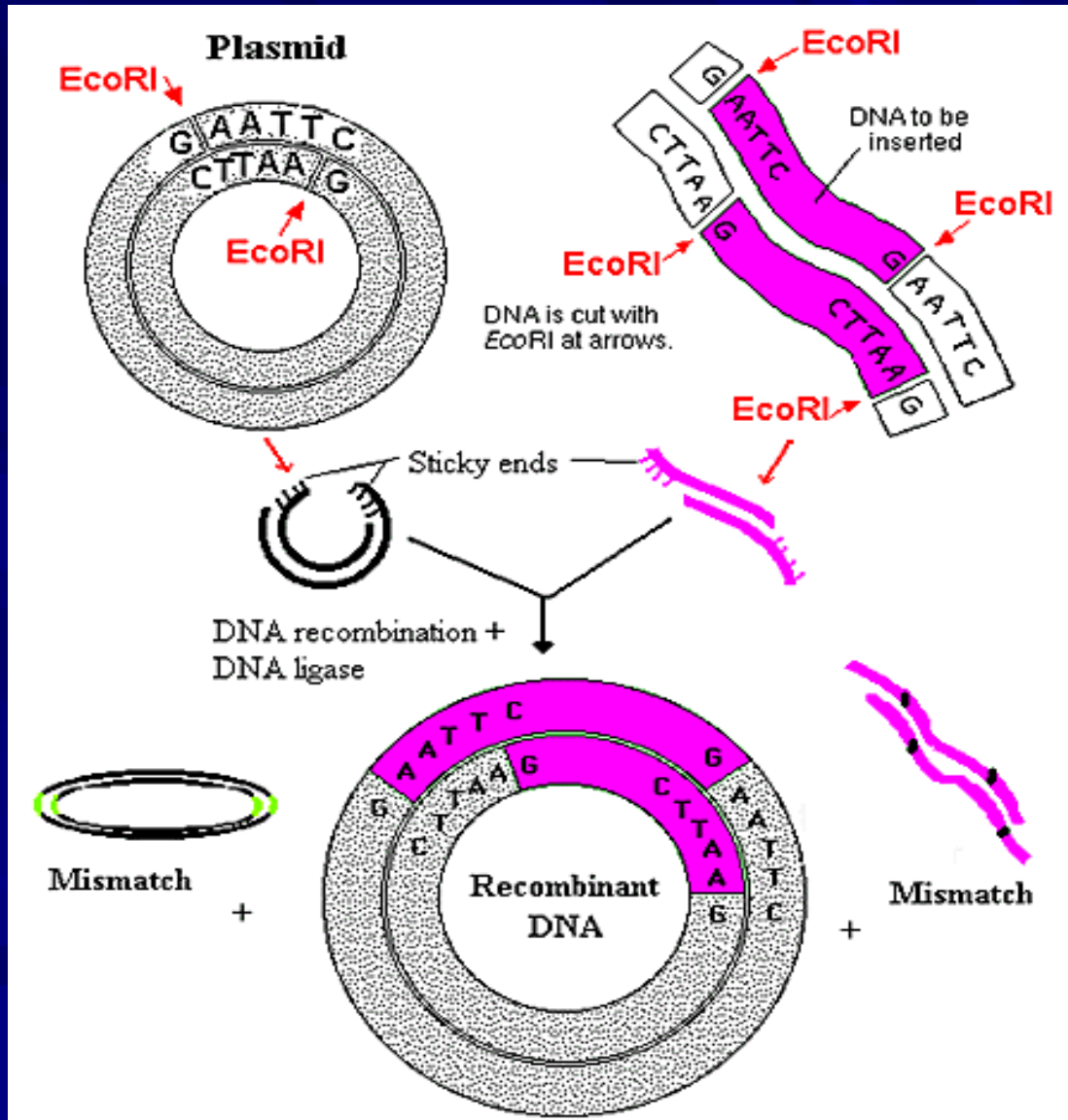
# DNA SEQUENCING



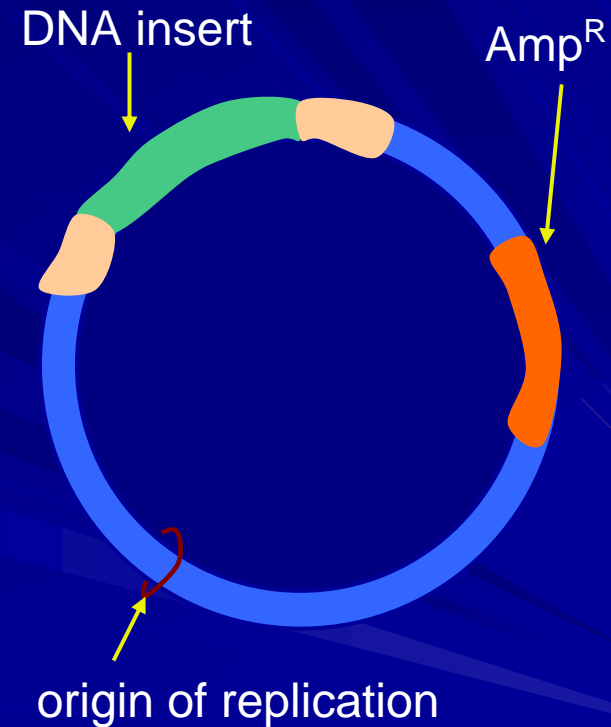
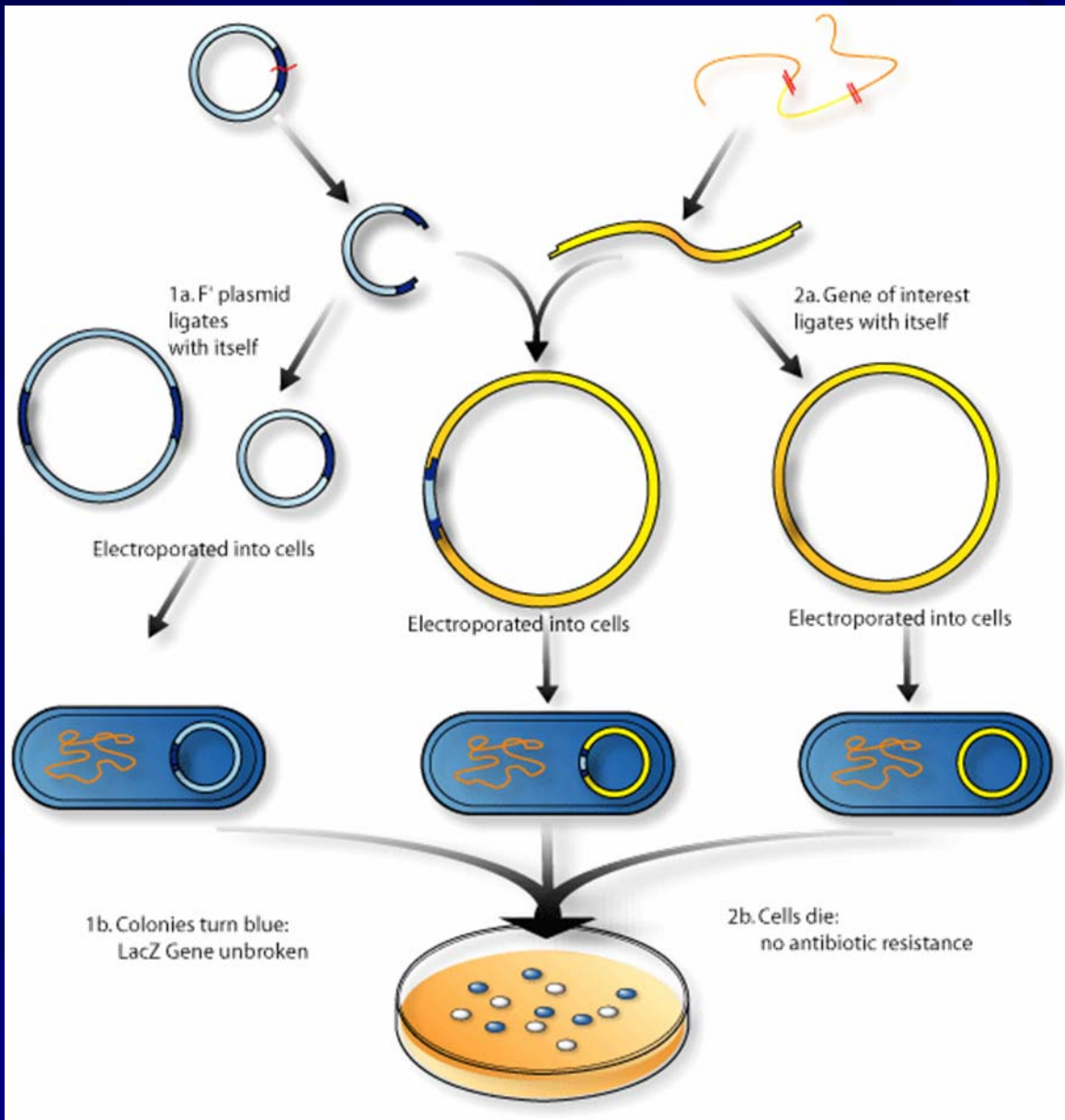
# DNA SEQUENCING

<http://www.dnalc.org/ddnalc/resources/cycseq.html>

# CLONING – PLASMID VECTOR



# CLONING – identifying transformed cells





# VECTORS

Vector	Form	Host	Carrying Capacity	Major Uses
Plasmid	Double-stranded circular DNA	<i>E. coli</i>	Upto 15 kb	cDNA libraries; subcloning
Bacteriophage lambda	Virus – linear DNA	<i>E. coli</i>	Upto 25 kb	Genomic and cDNA libraries
Cosmid	Double-stranded circular DNA	<i>E. coli</i>	30 – 45 kb	Genomic libraries
Bacteriophage P1	Virus – circular DNA	<i>E. coli</i>	70 – 90 kb	Genomic libraries
BAC	Bacterial artificial chromosome	<i>E. coli</i>	100 – 500 kb	Genomic libraries
YAC	Yeast artificial chromosome	Yeast	250 – 2000 kb	Genomic libraries



# GENOME SEQUENCING

**Genome sequencing:**

<http://www.pbs.org/wgbh/nova/genome/sequencer.html#>

**Whole genome shotgun sequencing:**

<http://smcg.cifn.unam.mx/enp-unam/03-EstructuraDelGenoma/animaciones/humanShot.swf>



# *What is bioinformatics?*

Research, development or application of computational tools and approaches to expand the use, acquisition, visualization, analysis, organization and archiving of biological, medical, behavioral or health data.

*[Bioinformatics at the NIH, 2001]*

<http://grants.nih.gov/grants/bistic/bistic.cfm>

# Important databases in the public domain

- National Center for Biotechnology Information (NCBI)

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

- European Bioinformatics Institute (<http://www.ebi.ac.uk/>)

- European Molecular Biology Laboratory (<http://www.embl.org>)

- DNA Data Bank of Japan (<http://www.ddbj.nig.ac.jp/Welcome.html>)

- TIGR (<http://www.tigr.org>)

# The National Center for Biotechnology Information (NCBI)



*Created in 1988 ( National Library of Medicine at NIH)*

- Establish public databases
- Conduct research in computational biology
- Develop software tools for sequence analysis
- Disseminate biomedical information

# NCBI database types

– Bibliographic



**Citations for biomedical articles**

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>



**Free archive of life sci. journals**

<http://www.pubmedcentral.nih.gov/>



# NCBI database types

– Bibliographic



**Books that can be searched online**

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books&itool=toolbar>



**Human genes/genetic disorders**

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

# NCBI database types

- Sequence (nucleotide; protein)
- Taxonomy
- Genome
- Gene
- Expression
- Structure

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



# Types of Sequence Databases

## Primary Databases

- **Contain raw and redundant data:** original experimental sequences, submitted and “owned” by experimentalists
- **Database staff review and organize the data:** don't add, modify or update the records

➤ *Examples:* GenBank, SNP, GEO

# Types of Sequence Databases

## Derivative Databases

- **Human-curated** (data compilation and correction)
  - *Examples:* LocusLink, OMIM & Literature databases
- **Computationally-Derived** (auto-partitioning GenBank seqs)
  - *Example:* UniGene
- **Combination**
  - *Examples:* RefSeq, Genome Assembly

# 1° Sequence Database

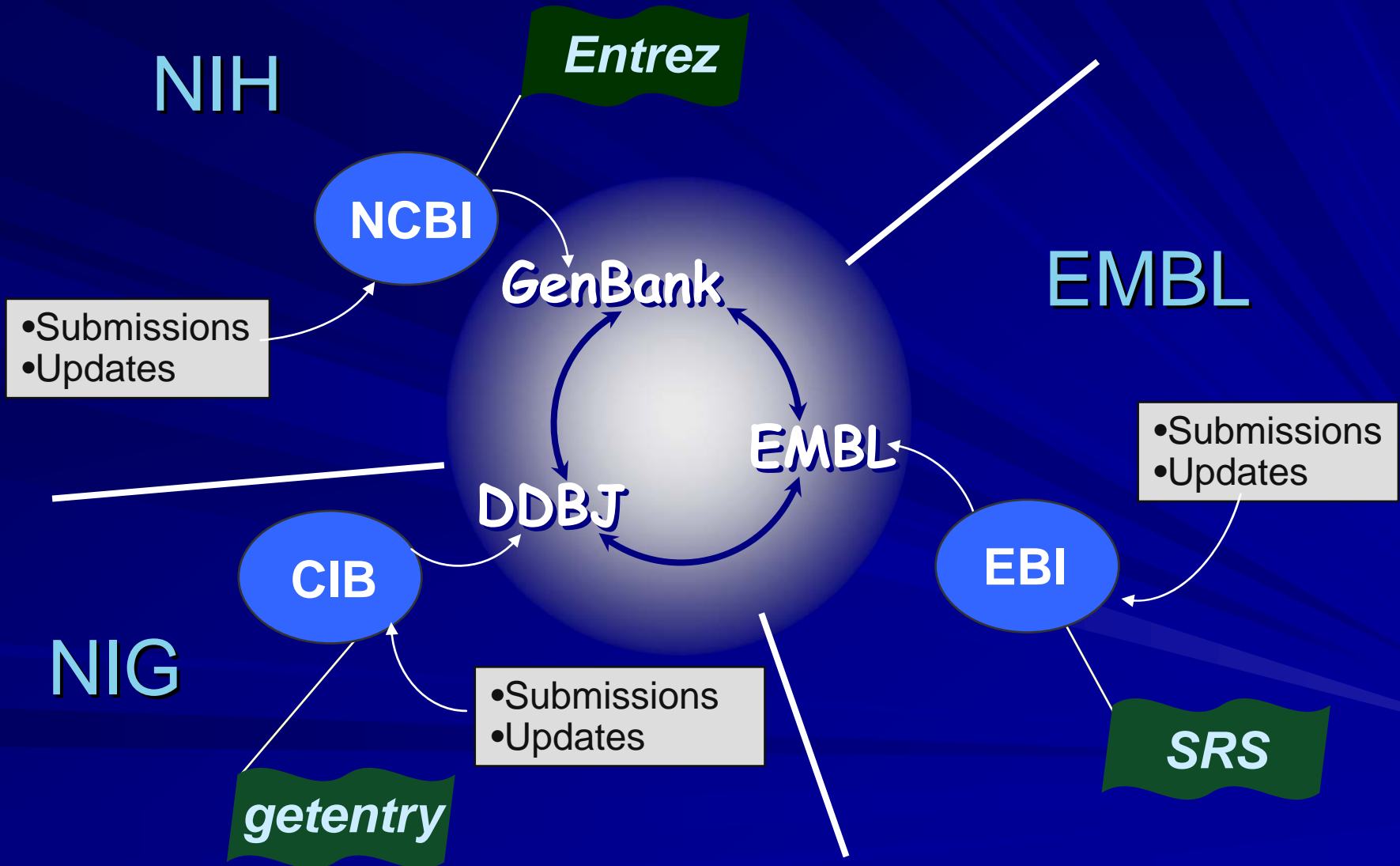
GenBank

- **Nucleotide-only sequence database**
- **Archival (>292,000 organisms)**

## Submission of GenBank Data to NCBI:

- Direct submissions of individual records via Web (*BankIt, Sequin*)
- Batch submissions of bulk sequences via e-mail (*EST, dbGSS, dbSTS*)
- FTP accounts for sequencing centers

# The International Sequence Database Collaboration



# Check for cross-functionality of accession numbers

Accession no. AB062786

EBI: <http://www.ebi.ac.uk>

DDBJ: <http://www.ddbj.nig.ac.jp/>

# Organization of GenBank: GenBank Divisions (gbdiv)

Records are divided into 18 divisions:

- 1 Patent
- 5 High Throughput
- 12 Traditional

## Traditional Divisions:

### Bulk Divisions:

- Batch Submission  
(Email and FTP)
- Inaccurate
- Poorly characterized

EST	Expressed Sequence Tag
PRI	Primate
PLN	Plant and Fungal
BCT	Bacterial and Archaeal
INV	Invertebrate
ROD	Rodent
VRL	Viral
VRT	Other Vertebrate
MAM	Mammalian (ex. ROD and PRI)
PHG	Phage
SYN	Synthetic (cloning vectors)
UNA	Unannotated
ENV	Environmental

**Length**

**mRNA = cDNA  
DNA = genomic**

**Division**

□ 1: [AB062786](#). Photinus pyralis ... [g... 51]

LOCUS AB062786 927 bp mRNA linear INV 27-SEP-2001

DEFINITION Photinus pyralis mRNA for luciferin regenerating enzyme, complete cds.

ACCESSION [AB062786](#) REGION: 1..927

VERSION AB062786.1 GI:14331151

**Accession Number**

KEYWORDS .

SOURCE Photinus pyralis (common eastern firefly)

**Accession.Version**

ORGANISM [Photinus pyralis](#)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Elateriformia; Cantharoidea; Lampyridae; Photinus.

**NCBI's Taxonomy**

REFERENCE 1

AUTHORS Gomi,K. and Kajiyama,N.

TITLE Oxyluciferin, a luminescence product of firefly luciferase, is enzymatically regenerated into luciferin

JOURNAL J. Biol. Chem. 276 (39), 36508-36513 (2001)

MEDLINE [21453302](#)

PUBMED [11457857](#)

REFERENCE 2 (bases 1 to 927)

AUTHORS Gomi,K. and Kajiyama,N.

TITLE Direct Submission

JOURNAL Submitted (07-JUN-2001) Keiko Gomi, Kikkoman Corporation, Research & Development; Noda 399, Noda-shi, Chiba 2780037, Japan (E-mail:8553@mail.kikkoman.co.jp, Tel:81471235579)

FEATURES

Location/Qualifiers

**Feature Table**

source

1..927  
 /organism="Photinus pyralis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7054"

CDS

1..927  
 /codon\_start=1  
 /product="luciferin regenerating enzyme"  
 /protein\_id="[BAB60700.1](#)"  
 /db\_xref="GI:14331152"

**GenPept Protein ID**

/translation="MGPVVEKIAELGKYTVGEGPHUDHETQTLYFVDTVEKTFHKYVP  
 SQKKYTFCKVDKLVSFIIPLAGSPGRFVWSLEREIAILTWGVSAAPTSIEAIVNVEP  
 HIKNNRLNDGKADPLGNLWTGTMAIDAGLPVGPVVTGSLYHLGADKKVKMHESNIAIAN  
 GLAWSNDLKKMYIDSGKRRVDEYDYDASTLSISNQRPLFTFEKHEVPGYPDGGQTIDE  
 EGNLWVAVFQGQRIKISTQQPEVLLDTPVKIPDPQVTSVAFGGPNLDELHVTSAGLQL  
 DDSSLDKSLVNGHVYRVTGLGVKGFAGVKVKL"

ORIGIN

```

1 atggggccag ttggtgaaaa aattgcagaa cttggcaagt atacggttgg agaaggtcct
61 cactgggatc atgaaactca gaccttatat ttcgtcgaca ccgtagagaa aacttttcat
121 aaatatgtac cttctcagaa aaaatacacg ttttgtaaag tagataaact ggtttcttte
181 attatceccc ttgctgggat ccttggecgt tttgtagtca gtttggaacg tgaaatagcc
241 attettacat gggatggcgt tagtgctgca cctacaagca tagaagctat tgттаатgтc
301 gaaccacaca ttaaaaataa cagactcaat gatggcaaag cagateccct tggcaatcta
361 tggacaggta caatggetat tgacgctggt ctccccgtag gaccggtcac tggcagttta
421 tatcatttag gggctgataa aaaggtaaaa atgcacgaga gcaacatagc tatagcaaat
481 gggctcgcgt ggagtaatga tttgaagaaa atgtattata ttgattcggg gaaaagaaga
541 gtagacgagt acgattatga tgettetaca ttatecatca gcaatcaacg gccattattt
601 acttttgaaa agcatgaagt gcctggatat ccagatggtc aaacaattga tgaggagggt
    
```





## National Center for Biotechnology Information

National Library of Medicine

National Institutes of Health

PubMed

Entrez

BLAST

OMIM

Books

TaxBrowser

Structure

Search



for

### SITE MAP

Guide to NCBI resources

### About NCBI

An introduction for researchers, educators and the public

### GenBank

Sequence submission support and software

### Literature databases

PubMed, OMIM, Books, and PubMed Central

### ▶ What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

### **HIV-1 Protein Interaction Database**

HIV/AIDS researchers can now access a database of known interactions of HIV-1 proteins with proteins from human hosts. The database offers a concise summary of these interactions with links to PubMed, sequence data, and genes. [Read more...](#)

### Hot Spots

- ▶ Clusters of orthologous groups
- ▶ Coffee Break, Genes & Disease, NCBI Handbook
- ▶ Electronic PCR
- ▶ Entrez Home
- ▶ Entrez Tools
- ▶ Gene expression omnibus (GEO)
- ▶ Human genome resources

## e.g. - pharmacogenetics

- Identifying novel targets for new drugs
  - mapping and identifying genes associated w/ disease
  - characterizing proteins ———> targets for new drugs
  
- Identifying genetic variants associated w/ adverse drug reactions
  - e.g., cytochrome P450s = multigene family of enzymes (liver)
  - genetically variable expression = variation in drug efficacy

# Potential consequences of polymorphic drug metabolism

- Extended pharmacological effect
- Adverse drug reactions
- Lack of pro-drug activation (e.g., codeine)
- Drug toxicity
- Increased effective dose
- Metabolism by alternate, deleterious pathways
- Exacerbated drug – drug interactions

## Common pharmacogenetic polymorphisms in human drug metabolizing enzymes (Weber, W.W. *Pharmacogenetics*. Oxford, 1997)

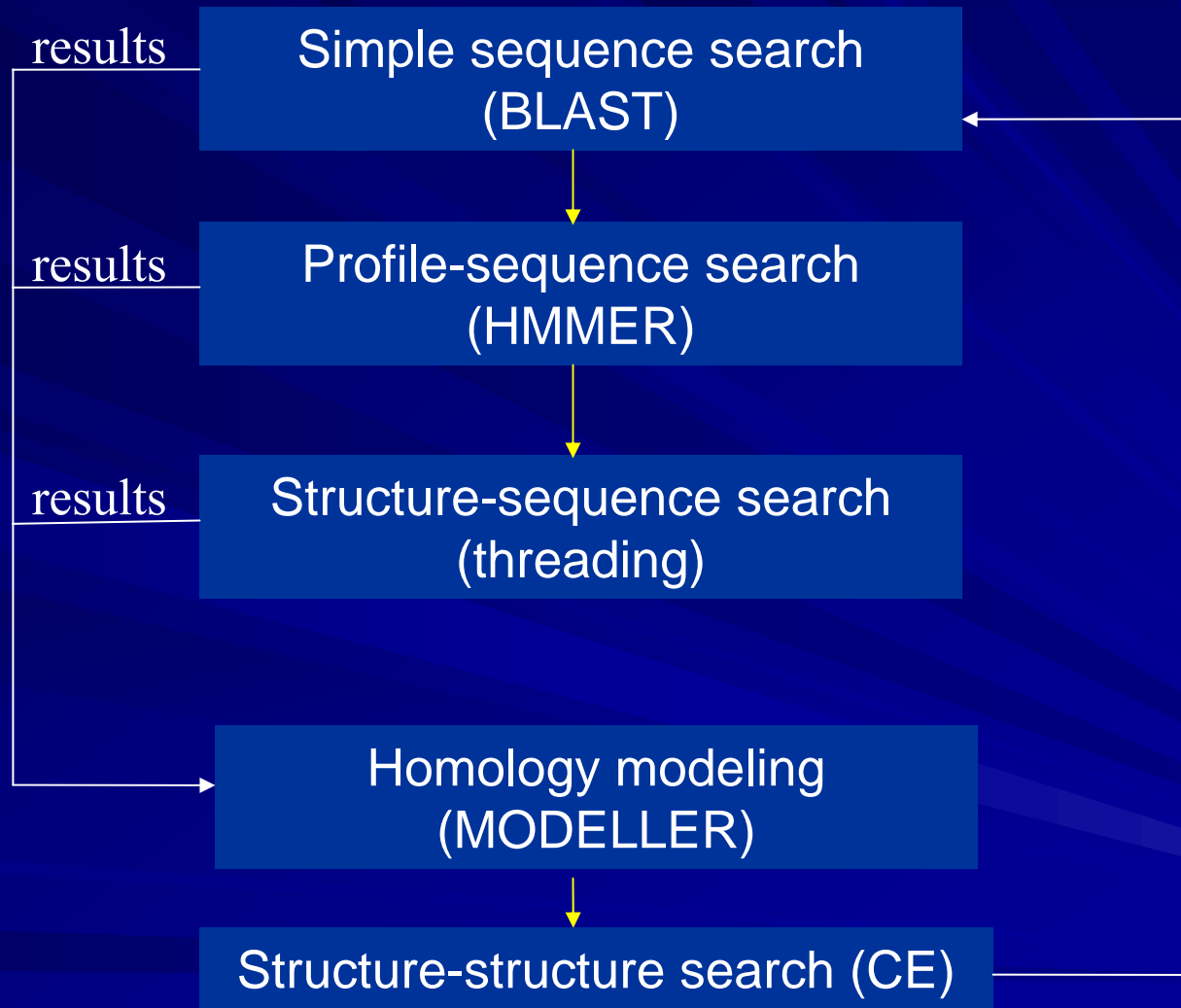
Gene	Metaboliser Phenotype	Frequency	# of drugs	Examples
CYP2D6	Poor	White 6%, African American 2%	>100	codeine, dextromethorphan
	Ultra-rapid	Ethiopian 20%, Spanish 7%		
CYP2C9	Reduced		>60	Ibuprofen, warfarin
TPMT	Poor	low in all populations	<10	6-mercaptopurine, 6-thioguanine

### Example: Cytochrome P450 gene - CYP2D6

- CYP2D6 is highly polymorphic (inactive in ~ 6% of Caucasians)
  - codes for debrisoquine hydroxylase

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

# Sequence/structure searching tools



# Tool comparison



**BLAST**

**HMM**

**Threading**

**Sensitivity:** Least sensitive  Most sensitive

**Speed:** Seconds  Minutes  Hours

**DB size:**  $1 \times 10^6$    $1 \times 10^6$   18000 (PDB)

**Result interpretation:** Relatively easy  Some expertise required

# Sequence similarity searching

## Why do it?

- identify and annotate sequences with no, incomplete, incorrect annotations (GenBank)
- infer functionality for genes/proteins
- find conserved domains
- assemble genomes; clean up sequences (e.g., suspected cloning vector sequences)
- explore evolutionary relationships

**NOTE:** Similar sequences may NOT be homologous!



# Basic Local Alignment Search Tool (BLAST)

- Calculates similarity for biological sequences
- Finds best local alignments
- Searches for matching “words” rather than individual residues
- Uses statistical theory to determine if a match might have occurred by chance

# Sequence Alignment

**Global alignment:** compare sequences over entire length  
(dynamic – e.g., Needleman-Wunsch)

- identify long insertions/deletions
- check data quality

**Local alignment:** compare segments of sequences  
(heuristic -- BLAST; FASTA; Smith-Waterman) <sup>*speed*</sup>

- high quality alignments

**Dot plot:** exploration of two entire sequences for similarity

- repeat discovery
- identify long insertions/deletions

# Basic Local Alignment Search Tool (BLAST)

<http://www.ncbi.nlm.nih.gov/BLAST/>



# BLAST

PubMed

Entrez

BLAST

OMIM

Taxonomy

Structure

Info

- FAQs
- News
- References
- NCBI Contributors

Education

- Program selection guide
- Tutorial
- URL API guide

Download

- Databases
- Documentation
- Executables
- Source code

**NEW 12 May 2004** BLAST 2.2.9 has been released. [Read more...](#)

## Nucleotide

- [Discontiguous megablast](#)
- [Megablast](#)
- [Nucleotide-nucleotide BLAST \(blastn\)](#)
- [Search for short, nearly exact matches](#)
- [Search trace archives with megablast or discontiguous megablast](#)

## Translated

- [Translated query vs. protein database \(blastx\)](#)
- [Protein query vs. translated database \(tblastn\)](#)
- [Translated query vs. translated database \(tblastx\)](#)

## Special

- [Search for gene expression data \(GEO BLAST\)](#)

## Protein

- [Protein-protein BLAST \(blastp\)](#)
- [PHI- and PSI-BLAST](#)
- [Search for short, nearly exact matches](#)
- [Search the conserved domain database \(rpsblast\)](#)
- [Search by domain architecture \(cdart\)](#)

## Genomes

- [Chicken, cow, pig, dog, sheep, cat](#) **NEW**
- [Environmental samples](#)
- [Human, mouse, rat](#)
- [Fugu rubripes, zebrafish](#)
- [Insects, nematodes, plants, fungi, malaria](#)
- [Microbial genomes, other eukaryotic genomes](#)

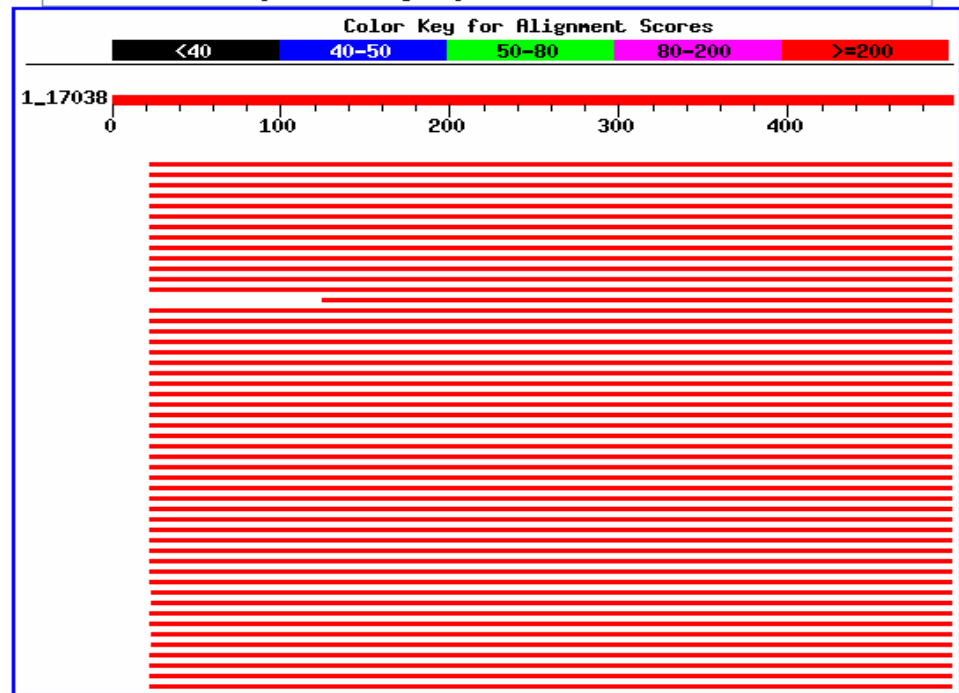
## Meta

- [Retrieve results by RID](#)

## Distribution of 509 Blast Hits on the Query Sequence

# Sample BLAST 1

BAA23125 CYP2D4 [Rattus norvegicus]..S= 722 E=0.0



### Related Structures

Sequences producing significant alignments:	Score (bits)	E Value	
<a href="#">gi 40805836 ref NP_000097.2 </a> cytochrome P450, subfamily IID...	<a href="#">933</a>	0.0	<a href="#">G</a>
<a href="#">gi 50960849 gb AAH75024.1 </a> Cytochrome P450, subfamily IID, ...	<a href="#">927</a>	0.0	<a href="#">G</a>
<a href="#">gi 117246 sp P10635 CP2D6</a> HUMAN Cytochrome P450 2D6 (CYPIID...	<a href="#">926</a>	0.0	<a href="#">G</a>
<a href="#">gi 18481430 gb AAL73443.1 </a> cytochrome P450 2D [Macaca fuscata]	<a href="#">895</a>	0.0	
<a href="#">gi 62088564 dbj BAD92729.1 </a> Debrisoquine 4-hydroxylase muta...	<a href="#">880</a>	0.0	
<a href="#">gi 29465680 gb AAL92448.1 </a> cytochrome P450 2D [Callithrix j...	<a href="#">879</a>	0.0	
<a href="#">gi 2493367 sp Q29488 CP2DH</a> MACFA Cytochrome P450 2D17 (CYPI...	<a href="#">869</a>	0.0	
<a href="#">gi 49066333 gb AAT49268.1 </a> cytochrome P450 CYP2D42 [Macaca ...	<a href="#">868</a>	0.0	
<a href="#">gi 50897272 ref NP_001002910.1 </a> cytochrome P450, family 2, ...	<a href="#">858</a>	0.0	<a href="#">G</a>
<a href="#">gi 3913340 sp O18992 CP2DJ</a> CALJA Cytochrome P450 2D19 (CYPI...	<a href="#">840</a>	0.0	
<a href="#">gi 181306 gb AAA35737.1 </a> debrisoquine 4-hydroxylase	<a href="#">810</a>	0.0	<a href="#">G</a>
<a href="#">gi 47678391 emb CAG30316.1 </a> CYP2D6 [Homo sapiens]	<a href="#">805</a>	0.0	<a href="#">G</a>
<a href="#">gi 57209874 emb CAI43003.1 </a> GD:CYP2D6 [Homo sapiens] >gi 45...	<a href="#">805</a>	0.0	<a href="#">G</a>
<a href="#">gi 522195 gb AAA36403.1 </a> cytochrome P450db1	<a href="#">753</a>	0.0	<a href="#">G</a>
<a href="#">gi 27806959 ref NP_776954.1 </a> cytochrome P450, subfamily IID...	<a href="#">728</a>	0.0	<a href="#">G</a>
<a href="#">gi 294 emb CAA48149.1 </a> cytochrome P-450IID [Bos taurus] >gi...	<a href="#">728</a>	0.0	<a href="#">G</a>

<a href="#">gi 3878450 emb CAA90617.1 </a>	Hypothetical protein K09A11.4 [C...	<a href="#">235</a>	2e-60	
<a href="#">gi 3913344 sp O73853 CP17A ICTPU</a>	Cytochrome P450 17A1 (CYPX...	<a href="#">235</a>	2e-60	
<a href="#">gi 21703936 ref NP_663449.1 </a>	similar to nuclear protein, 25...	<a href="#">235</a>	3e-60	
<a href="#">gi 13592296 gb AAK31395.1 </a>	Cytochrome p450 family protein 2...	<a href="#">234</a>	3e-60	
<a href="#">gi 7430625 pir  T15318</a>	hypothetical protein B0304.3 - Caeno...	<a href="#">234</a>	3e-60	
<a href="#">gi 3913305 sp P56590 CP1A1 CANFA</a>	Cytochrome P450 1A1 (CYPIA...	<a href="#">234</a>	6e-60	
<a href="#">gi 57108669 ref XP_544773.1 </a>	PREDICTED: similar to cytochro...	<a href="#">234</a>	6e-60	
<a href="#">gi 39587860 emb CAE67878.1 </a>	Hypothetical protein CBG13471 [...	<a href="#">233</a>	1e-59	
<a href="#">gi 30341 emb CAA26458.1 </a>	cytochrome P(1)-450 [Homo sapiens]	<a href="#">233</a>	1e-59	
<a href="#">gi 61364554 gb AAx42562.1 </a>	cytochrome P450 family 1 subfami...	<a href="#">232</a>	2e-59	
<a href="#">gi 41059946 emb CAF18541.1 </a>	cytochrome P450 1A1 [Phoca groe...	<a href="#">232</a>	2e-59	
<a href="#">gi 181276 gb AAA52139.1 </a>	cytochrome P-450-1	<a href="#">232</a>	2e-59	

**Alignments**

[>gi|40805836|ref|NP\\_000097.2|](#) cytochrome P450, subfamily IID, polypeptide 6 [Homo sapiens]  
[gi|57209873|emb|CAI43002.1|](#) GD:CYP2D6 [Homo sapiens]  
[gi|45768272|gb|AAH67432.1|](#) Cytochrome P450, subfamily IID, polypeptide 6 [Homo sapiens]  
 Length = 497

Score = 933 bits (2411), Expect = 0.0  
 Identities = 461/475 (97%), Positives = 461/475 (97%)

```

Query: 23  MHRRQRWAARYXXXXXXXXXXXXXXXXXHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLN 82
          MHRRQRWAARY                      HVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLN
Sbjct: 23  MHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLN 82

Query: 83  GLAAVREALVTHGEDTADRPPVPIQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNL 142
          GLAAVREALVTHGEDTADRPPVPIQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNL
Sbjct: 83  GLAAVREALVTHGEDTADRPPVPIQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNL 142

Query: 143 GLGKKSLEQWVTEEAACLCAAFANHSGRPFPRPNGLLDKAVSNVIAASLTCGRRFEYDDPRF 202
          GLGKKSLEQWVTEEAACLCAAFANHSGRPFPRPNGLLDKAVSNVIAASLTCGRRFEYDDPRF
Sbjct: 143 GLGKKSLEQWVTEEAACLCAAFANHSGRPFPRPNGLLDKAVSNVIAASLTCGRRFEYDDPRF 202

Query: 203 LRLDLLAQEGLKEESGFLREVLNAVVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTW 262
          LRLDLLAQEGLKEESGFLREVLNAVVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTW
Sbjct: 203 LRLDLLAQEGLKEESGFLREVLNAVVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTW 262

Query: 263 DPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLCIVVADLFSAGMVTSTTLAWGLLLMI 322
          DPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLCIVVADLFSAGMVTSTTLAWGLLLMI
Sbjct: 263 DPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLCIVVADLFSAGMVTSTTLAWGLLLMI 322

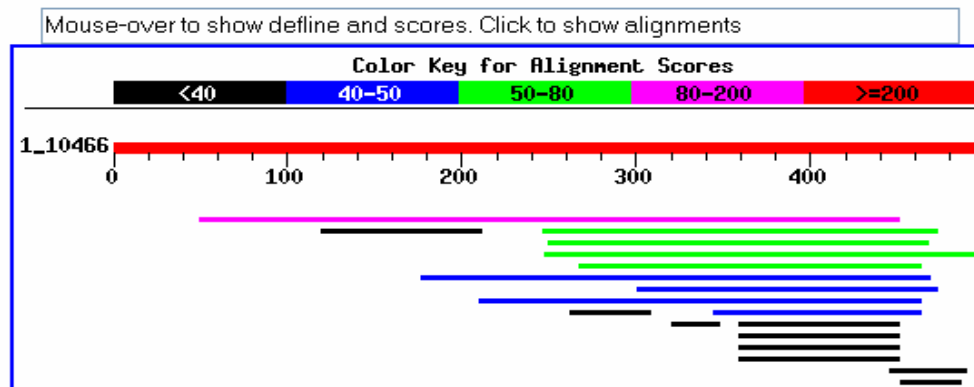
Query: 323 LHPDVQRRVQOEIDDVIGQVRRPEMGDQAHPYTTAVIHEVQRFQGDIVPLGVTHMTRSRI 382
          LHPDVQRRVQOEIDDVIGQVRRPEMGDQAHPYTTAVIHEVQRFQGDIVPLGVTHMTRSRI
Sbjct: 323 LHPDVQRRVQOEIDDVIGQVRRPEMGDQAHPYTTAVIHEVQRFQGDIVPLGVTHMTRSRI 382

Query: 383 EVQGFRIKGTTLITNLSSVLKDEAVWEKPFRRFHPEHFLDAQGHFVKPEAFLPFSAGRRA 442
          EVQGFRIKGTTLITNLSSVLKDEAVWEKPFRRFHPEHFLDAQGHFVKPEAFLPFSAGRRA
Sbjct: 383 EVQGFRIKGTTLITNLSSVLKDEAVWEKPFRRFHPEHFLDAQGHFVKPEAFLPFSAGRRA 442
  
```

# Sample BLAST 1

## Distribution of 18 Blast Hits on the Query Sequence

# Sample BLAST 2: “cytochrome AND Archaea”



### Related Structures

Sequences producing significant alignments:

	Score (bits)	E Value	
<a href="#">gi 55229734 gb AAV45153.1 </a> cytochrome P450 [Haloarcula mari...	<a href="#">87</a>	2e-19	<a href="#">G</a>
<a href="#">gi 55232584 gb AAV48003.1 </a> cytochrome P450 [Haloarcula mari...	<a href="#">76</a>	5e-16	<a href="#">G</a>
<a href="#">gi 55230615 gb AAV46034.1 </a> cytochrome P450 [Haloarcula mari...	<a href="#">76</a>	5e-16	<a href="#">G</a>
<a href="#">gi 15789667 ref NP_279491.1 </a> Cyc [Halobacterium sp. NRC-1] ...	<a href="#">65</a>	7e-13	<a href="#">G</a>
<a href="#">gi 48839941 ref ZP_00296870.1 </a> COG2124: Cytochrome P450 [Me...	<a href="#">65</a>	7e-13	
<a href="#">gi 48852346 ref ZP_00306534.1 </a> COG2124: Cytochrome P450 [Fe...	<a href="#">48</a>	1e-07	
<a href="#">gi 48477157 ref YP_022863.1 </a> cytochrome P450 [Picropophilus t...	<a href="#">47</a>	3e-07	<a href="#">G</a>
<a href="#">gi 48852877 ref ZP_00307059.1 </a> COG2124: Cytochrome P450 [Fe...	<a href="#">45</a>	1e-06	
<a href="#">gi 48478471 ref YP_024177.1 </a> cytochrome P450 [Picropophilus t...	<a href="#">44</a>	2e-06	<a href="#">G</a>
<a href="#">gi 15921406 ref NP_377075.1 </a> hypothetical cytochrome P450 [...]	<a href="#">39</a>	7e-05	<a href="#">G</a>
<a href="#">gi 51247807 pdb 1UE8 A</a> Chain A, Crystal Structure Of Thermo...	<a href="#">39</a>	7e-05	<a href="#">S</a>
<a href="#">gi 7427578 pir  S71328</a> cytochrome P450 CYP119 - Sulfolobus ...	<a href="#">33</a>	0.003	<a href="#">S</a>
<a href="#">gi 13399503 pdb 1IO8 B</a> Chain B, Thermophilic Cytochrome P45...	<a href="#">33</a>	0.003	<a href="#">S</a>
<a href="#">gi 55230305 gb AAV45724.1 </a> cytochrome b(C-terminal)/b6 [Hal...	<a href="#">25</a>	1.4	<a href="#">G</a>
<a href="#">gi 2822447 gb AAC82953.1 </a> CydB [Halobacterium sp. NRC-1] >g...	<a href="#">24</a>	1.9	<a href="#">G</a>
<a href="#">gi 15789855 ref NP_279679.1 </a> CoxB1 [Halobacterium sp. NRC-1...	<a href="#">24</a>	2.4	<a href="#">G</a>
<a href="#">gi 18977077 ref NP_578434.1 </a> cytochrome c-type biogenesis p...	<a href="#">22</a>	7.1	<a href="#">G</a>
<a href="#">gi 13541204 ref NP_110892.1 </a> Cytochrome b subunit of the bc...	<a href="#">22</a>	9.3	<a href="#">G</a>

**Alignments**

Get selected sequences

Select all

Deselect all

# What is BLAST?

AATTGGCTAGCTAA

| | | | | | | |

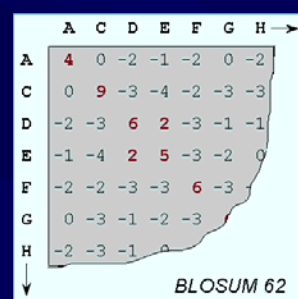
...AAAATGCAAA**ATGCGGGTAGCTT**ATTCTAGAAGATT...

**Matches: 10**

**Mismatches: 4**

Similarity score based on matches, mis-matches, gaps

# BLAST: Substitution Matrix and Gap Cost



A 8x8 substitution matrix for BLOSUM 62. The rows and columns are labeled A, C, D, E, F, G, H. The matrix values are: A: [4, 0, -2, -1, -2, 0, -2]; C: [0, 9, -3, -4, -2, -3, -3]; D: [-2, -3, 6, 2, -3, -1, -1]; E: [-1, -4, 2, 5, -3, -2, 0]; F: [-2, -2, -3, -3, 6, -3]; G: [0, -3, -1, -2, -3]; H: [-2, -3, -1]. The matrix is symmetric. The text 'BLOSUM 62' is at the bottom.

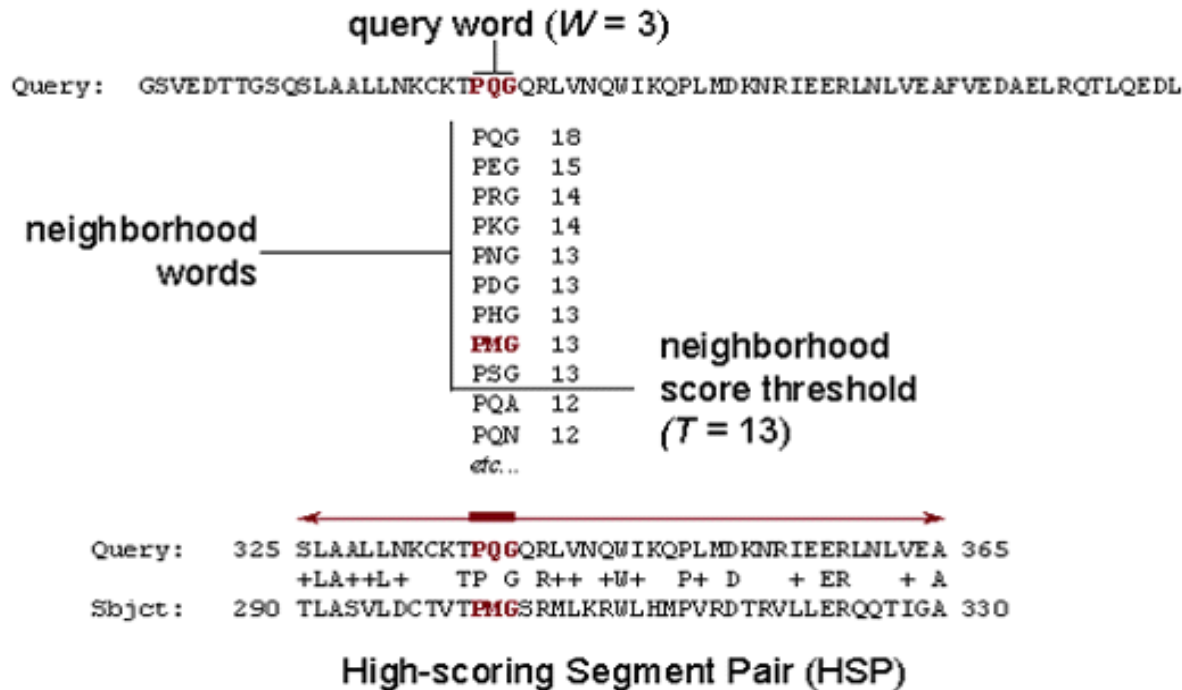
Query Length	Substitution Matrix	Gap cost
<35	PAM-30*	(9,1)
35-50	PAM-70	(10, 1)
50-85	BLOSUM-80	(10, 1)
>85	BLOSUM-62	(11, 1)

\*PAM = Percent Accepted Mutation; 1 PAM unit = 1% of aa in protein changed

- **BLOSUM-62 generally performs better than PAM**
- **PAM better if looking for distant relationships**



# The BLAST Search Algorithm



NCBI BLAST

- matrix used to create look-up tables of neighborhood words
- seeks pairs of similar segments whose score exceeds **threshold** (HSPs)
  - $T < 13$  not reported
  - locates “seeds” of similarity along query
  - extends seeds in both directions until max. possible score reached

# Protein Words

Query: **GTQITVEDLFFYNIATRRKALKN**

Word size = 3

**GTQ**

**TQI**

**QIT**

**ITV**

**TVE**

**VED**

**EDL...**

Word size 2 or 3 (default = 3)

W = 2; T = 16

W = 3; T = 32

Neighborhood Words

**ITV** → LTV, MTV, ISV, LSV, etc.

Make a lookup  
table of words

# Nucleotide Words

Query: **GTACTGGACATGGACCCTACAGGAA**

Word size = 11

Minimum word size = 7

**GTACTGGACAT**

blastn default = 11

**TACTGGACATG**

megablast default = 28

**ACTGGACATGG**

**CTGGACATGGA**

**TGGACATGGAC**

**GGACATGGACC**

**GACATGGACCC...**

Make a lookup  
table of words

## BLAST: Bit Score

Bit Score ( $S'$ ) : normalized raw score ( $S$ ), allows direct comparison of searches from diverse dbs

$$S' = (\lambda S - \ln K) / \ln 2$$

$S$  = raw score (sum of scores in substitution)

$K$  = variable; value dependent on matrix used

$\lambda$  = parameter used as natural scale for scoring system

# BLAST Statistics: E-value

E-value (E) : measure of statistical significance

e.g.,  $E=0.01 \rightarrow 1\%$  chance that match is due to a random event; dependent on db size

$$E = Kmne^{-\lambda S}$$

$K$  = variable; value dependent on matrix used

$m$  = length of query (nucleotide or aa)

$n$  = size of db

$\lambda$  = parameter used as natural scale for scoring system

$S$  = raw score (sum of scores in substitution)

# Tools for 3-D Structure Display and Searching

**Cn3D:** 3-D structure and sequence alignment viewer

--NCBI “Structure” db

**Domain Architecture Retrieval Tool (DART):**

--displays functional domains that make up a protein











--lists proteins with similar domain architectures

































**Vector Alignment Search Tool (VAST):**

--structure-structure similarity search program

**Threading:** algorithms for recognition of protein folding

## Welcome to the new Entrez cross-database search page

 <b>PubMed:</b> biomedical literature citations and abstracts 	 <b>Books:</b> online books 
 <b>PubMed Central:</b> free, full text journal articles 	 <b>OMIM:</b> Online Mendelian Inheritance in Man 
	 <b>Site Search:</b> NCBI web and FTP sites 

 <b>Nucleotide:</b> sequence database (GenBank) 	 <b>UniGene:</b> gene-oriented clusters of transcript sequences 
 <b>Protein:</b> sequence database 	 <b>CDD:</b> conserved protein domain database 
 <b>Genome:</b> whole genome sequences 	 <b>3D Domains:</b> domains from Entrez Structure 
 <b>Structure:</b> three-dimensional macromolecular structures 	 <b>UniSTS:</b> markers and mapping data 
 <b>Taxonomy:</b> organisms in GenBank 	 <b>PopSet:</b> population study data sets 
 <b>SNP:</b> single nucleotide polymorphism 	 <b>GEO Profiles:</b> expression and molecular abundance profiles 
 <b>Gene:</b> gene-centered information 	 <b>GEO DataSets:</b> experimental sets of GEO data 
 <b>HomoloGene:</b> Eukaryotic homology groups 	 <b>Cancer Chromosomes:</b> cytogenetic databases 

 <b>Journals:</b> detailed information <i>about</i> the journals indexed in PubMed and other 	 <b>MeSH:</b> detailed information about NLM's controlled vocabulary 
--	--

**Thank you!**

**Dr. Medha Devare**

[mhd6@cornell.edu](mailto:mhd6@cornell.edu)

**Life Sciences/Bioinformatics Specialist**

**Albert R. Mann Library**

**Cornell University, Ithaca, NY 14853**



# Exercise: BLASTp

Sequences producing significant alignments:		(Bits)	Value
<a href="#">gb AAT78551.1 </a>	thiopurine methyltransferase [Pseudomonas citrone	<a href="#">136</a>	6e-31
<a href="#">gb AAT49760.1 </a>	PA2832 [synthetic construct]	<a href="#">121</a>	2e-26
<a href="#">ref NP_251522.1 </a>	thiopurine methyltransferase [Pseudomonas ae...	<a href="#">121</a>	2e-26
<a href="#">ref ZP_01366300.1 </a>	hypothetical protein PaerPA_01003444 [Pseu...	<a href="#">118</a>	2e-25
<a href="#">ref ZP_00136158.2 </a>	COG0500: SAM-dependent methyltransferases ...	<a href="#">115</a>	1e-24
<a href="#">ref ZP_00974567.1 </a>	COG0500: SAM-dependent methyltransferases ...	<a href="#">113</a>	6e-24
<a href="#">ref ZP_01293671.1 </a>	hypothetical protein PaerP_01004452 [Pseud...	<a href="#">104</a>	2e-21
<a href="#">gb AAT78549.1 </a>	thiopurine methyltransferase [Pseudomonas aspleni	<a href="#">73.9</a>	4e-12
<a href="#">ref ZP_00417432.1 </a>	Thiopurine S-methyltransferase [Azotobacte...	<a href="#">73.2</a>	7e-12
<a href="#">gb AAT78509.1 </a>	thiopurine methyltransferase [uncultured bacte...	<a href="#">72.8</a>	9e-12
<a href="#">gb AAP12368.1 </a>	thiopurine methyltransferase [Pseudomonas sp. Hsa	<a href="#">72.0</a>	1e-11
<a href="#">gb ABC95207.1 </a>	TpmI [uncultured bacterium] >gb ABC95208.1  Tp...	<a href="#">72.0</a>	1e-11
<a href="#">gb ABC95213.1 </a>	TpmI [uncultured bacterium]	<a href="#">72.0</a>	1e-11
<a href="#">gb ABC95205.1 </a>	TpmH [uncultured bacterium]	<a href="#">72.0</a>	1e-11
<a href="#">ref YP_261341.1 </a>	thiopurine s-methyltransferase [Pseudomonas ...	<a href="#">71.2</a>	2e-11
<a href="#">ref NP_744025.1 </a>	thiopurine s-methyltransferase [Pseudomonas ...	<a href="#">68.9</a>	1e-10
<a href="#">ref ZP_00900176.1 </a>	thiopurine s-methyltransferase [Pseudomona...	<a href="#">68.9</a>	1e-10
<a href="#">ref YP_349713.1 </a>	Thiopurine S-methyltransferase [Pseudomonas ...	<a href="#">68.2</a>	2e-10
<a href="#">ref NP_879683.1 </a>	thiopurine S-methyltransferase [Bordetella p...	<a href="#">68.2</a>	2e-10
<a href="#">gb AAT78543.1 </a>	thiopurine methyltransferase [Pseudomonas pseu...	<a href="#">67.4</a>	4e-10
<a href="#">gb AAT78523.1 </a>	thiopurine methyltransferase [uncultured bacte...	<a href="#">67.4</a>	4e-10
<a href="#">gb AAT78513.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">67.0</a>	5e-10
<a href="#">gb AAT78500.1 </a>	thiopurine methyltransferase [uncultured bacte...	<a href="#">66.6</a>	6e-10
<a href="#">gb AAT78542.1 </a>	thiopurine methyltransferase [Halomonas halophila	<a href="#">66.6</a>	6e-10
<a href="#">gb AAT78546.1 </a>	thiopurine methyltransferase [Pseudomonas fluores	<a href="#">66.6</a>	6e-10
<a href="#">gb AAT78522.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">65.9</a>	1e-09
<a href="#">ref NP_884416.1 </a>	thiopurine S-methyltransferase [Bordetella p...	<a href="#">65.5</a>	1e-09
<a href="#">gb AAT78515.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">65.1</a>	2e-09
<a href="#">gb AAT78548.1 </a>	thiopurine methyltransferase [Pseudomonas chloror	<a href="#">64.7</a>	2e-09
<a href="#">gb AAT78547.1 </a>	thiopurine methyltransferase [Pseudomonas fragi]	<a href="#">64.3</a>	3e-09
<a href="#">gb AAT78545.1 </a>	thiopurine methyltransferase [Pseudomonas stutzer	<a href="#">64.3</a>	3e-09
<a href="#">gb ABC95216.1 </a>	TpmK [uncultured bacterium] >gb ABC95217.1  Tp...	<a href="#">64.3</a>	3e-09
<a href="#">gb AAT78521.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">63.9</a>	4e-09
<a href="#">ref YP_607250.1 </a>	thiopurine s-methyltransferase [Pseudomonas ...	<a href="#">62.4</a>	1e-08
<a href="#">ref YP_392024.1 </a>	Thiopurine S-methyltransferase [Thiomicrospi...	<a href="#">62.4</a>	1e-08
<a href="#">gb ABC95206.1 </a>	TpmL [uncultured bacterium]	<a href="#">62.0</a>	2e-08
<a href="#">gb AAT78550.1 </a>	thiopurine methyltransferase [Pseudomonas viridif	<a href="#">61.6</a>	2e-08
<a href="#">gb ABC95203.1 </a>	TpmA [uncultured bacterium]	<a href="#">60.5</a>	4e-08
<a href="#">gb AAT78526.1 </a>	thiopurine methyltransferase [uncultured bacte...	<a href="#">60.1</a>	6e-08
<a href="#">gb AAT78520.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">60.1</a>	6e-08
<a href="#">gb AAT78554.1 </a>	thiopurine methyltransferase [Pseudomonas syri...	<a href="#">58.9</a>	1e-07
<a href="#">gb AAT78531.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">58.9</a>	1e-07
<a href="#">ref NP_791602.1 </a>	thiopurine s-methyltransferase [Pseudomonas ...	<a href="#">58.9</a>	1e-07
<a href="#">ref ZP_00606326.1 </a>	Thiopurine S-methyltransferase [Magnetococ...	<a href="#">58.9</a>	1e-07
<a href="#">gb AAT78532.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">58.5</a>	2e-07
<a href="#">gb AAT78514.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">58.5</a>	2e-07
<a href="#">gb ABI60489.1 </a>	Thiopurine S-methyltransferase [Nitrosomonas eutr	<a href="#">57.0</a>	5e-07
<a href="#">gb AAT78525.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">56.2</a>	8e-07

# Exercise: BLASTp continued

<a href="#">gb AAT78557.1 </a>	thiopurine methyltransferase [Pseudomonas syri...	<a href="#">54.7</a>	2e-06
<a href="#">gb AAT78559.1 </a>	thiopurine methyltransferase [Pseudomonas syri...	<a href="#">54.7</a>	2e-06
<a href="#">ref YP_432621.1 </a>	SAM-dependent methyltransferase [Hahella che...	<a href="#">54.7</a>	2e-06
<a href="#">ref YP_236687.1 </a>	Thiopurine S-methyltransferase [Pseudomonas ...	<a href="#">54.7</a>	2e-06
<a href="#">gb AAT78562.1 </a>	thiopurine methyltransferase [Pseudomonas syri...	<a href="#">53.9</a>	4e-06
<a href="#">pdb 1PJZ A</a>	Chain A, Solution Structure Of Thiopurine Methyltr...	<a href="#">53.9</a>	4e-06
<a href="#">sp O86262 TPMT PSESJ</a>	Thiopurine S-methyltransferase (Thiopuri...	<a href="#">53.9</a>	4e-06
<a href="#">ref ZP_01074369.1 </a>	thiopurine methyltransferase [Marinomonas ...	<a href="#">53.9</a>	4e-06
<a href="#">ref ZP_01037995.1 </a>	thiopurine S-methyltransferase family prot...	<a href="#">53.9</a>	4e-06
<a href="#">gb AAT78558.1 </a>	thiopurine methyltransferase [Pseudomonas syri...	<a href="#">53.5</a>	5e-06
<a href="#">gb AAT78483.1 </a>	thiopurine methyltransferase [uncultured bacte...	<a href="#">53.5</a>	5e-06
<a href="#">ref ZP_01127750.1 </a>	Thiopurine S-methyltransferase [Nitrococcu...	<a href="#">53.5</a>	5e-06
<a href="#">ref ZP_01110366.1 </a>	thiopurine S-methyltransferase family prot...	<a href="#">52.4</a>	1e-05
<a href="#">emb CAC44176.3 </a>	probable thiopurine methyltransferase [Pseudomon...	<a href="#">52.0</a>	2e-05
<a href="#">gb AAT78555.1 </a>	thiopurine methyltransferase [Pseudomonas syri...	<a href="#">52.0</a>	2e-05
<a href="#">sp Q93JT2 TPMT PSEST</a>	Thiopurine S-methyltransferase (Thiopurine	<a href="#">52.0</a>	2e-05
<a href="#">ref ZP_00838694.1 </a>	Thiopurine S-methyltransferase [Shewanella...	<a href="#">51.2</a>	3e-05
<a href="#">ref ZP_00585834.1 </a>	Thiopurine S-methyltransferase [Shewanella...	<a href="#">51.2</a>	3e-05
<a href="#">gb AAT78519.1 </a>	thiopurine methyltransferase [uncultured bacteriu	<a href="#">50.8</a>	3e-05
<a href="#">ref ZP_01166521.1 </a>	Thiopurine S-methyltransferase [Oceanospir...	<a href="#">50.8</a>	3e-05
<a href="#">ref ZP_01198987.1 </a>	thiopurine S-methyltransferase [Xanthobact...	<a href="#">50.4</a>	5e-05
<a href="#">ref YP_363227.1 </a>	thiopurine S-methyltransferase [Xanthomonas ...	<a href="#">50.1</a>	6e-05
<a href="#">emb CAJ48627.1 </a>	thiopurine S-methyltransferase [Bordetella avium	<a href="#">49.7</a>	8e-05
<a href="#">ref YP_283741.1 </a>	Thiopurine S-methyltransferase [Dechloromona...	<a href="#">49.7</a>	8e-05
<a href="#">ref YP_732718.1 </a>	Thiopurine S-methyltransferase [Shewanella s...	<a href="#">49.3</a>	1e-04
<a href="#">ref YP_739487.1 </a>	Thiopurine S-methyltransferase [Shewanella s...	<a href="#">49.3</a>	1e-04
<a href="#">ref ZP_00849118.1 </a>	Thiopurine S-methyltransferase [Shewanella...	<a href="#">48.9</a>	1e-04
<a href="#">ref NP_641774.1 </a>	thiopurine methyltransferase [Xanthomonas ax...	<a href="#">48.5</a>	2e-04
<a href="#">ref YP_412943.1 </a>	Thiopurine S-methyltransferase [Nitrosospira...	<a href="#">48.1</a>	2e-04
<a href="#">ref ZP_00749132.1 </a>	COG0500: SAM-dependent methyltransferases [Vi	<a href="#">47.8</a>	3e-04
<a href="#">gb ABC95209.1 </a>	TpmAA [uncultured bacterium]	<a href="#">47.4</a>	4e-04
<a href="#">ref NP_716217.1 </a>	thiopurine S-methyltransferase [Shewanella o...	<a href="#">47.4</a>	4e-04
<a href="#">ref YP_191113.1 </a>	Thiopurine S-methyltransferase [Gluconobacte...	<a href="#">47.0</a>	5e-04
<a href="#">ref ZP_00756058.1 </a>	COG0500: SAM-dependent methyltransferases ...	<a href="#">47.0</a>	5e-04
<a href="#">ref ZP_00751977.1 </a>	COG0500: SAM-dependent methyltransferases ...	<a href="#">47.0</a>	5e-04
<a href="#">ref NP_230871.1 </a>	thiopurine methyltransferase [Vibrio cholera...	<a href="#">47.0</a>	5e-04
<a href="#">gb AAR37654.1 </a>	thiopurine S-methyltransferase [uncultured bacter	<a href="#">46.6</a>	7e-04
<a href="#">ref ZP_01223026.1 </a>	Putative thiopurine methyltransferase [Pho...	<a href="#">45.8</a>	0.001
<a href="#">ref YP_129893.1 </a>	Putative thiopurine methyltransferase [Photo...	<a href="#">45.8</a>	0.001
<a href="#">ref ZP_00581333.1 </a>	Thiopurine S-methyltransferase [Shewanella...	<a href="#">45.8</a>	0.001
<a href="#">ref YP_269794.1 </a>	thiopurine S-methyltransferase family protei...	<a href="#">45.4</a>	0.001
<a href="#">ref ZP_01065138.1 </a>	thiopurine methyltransferase [Vibrio sp. M...	<a href="#">45.4</a>	0.001
<a href="#">ref ZP_00993235.1 </a>	thiopurine methyltransferase [Vibrio splen...	<a href="#">45.4</a>	0.001
<a href="#">ref XP_503586.1 </a>	hypothetical protein [Yarrowia lipolytica] >...	<a href="#">45.1</a>	0.002
<a href="#">ref NP_636768.1 </a>	thiopurine methyltransferase [Xanthomonas ca...	<a href="#">45.1</a>	0.002
<a href="#">ref NP_797675.1 </a>	thiopurine methyltransferase [Vibrio parahae...	<a href="#">45.1</a>	0.002
<a href="#">gb ABC95212.1 </a>	TpmAB [uncultured bacterium]	<a href="#">44.7</a>	0.002
<a href="#">ref ZP_01433693.1 </a>	Thiopurine S-methyltransferase [Shewanella...	<a href="#">43.9</a>	0.004
<a href="#">gb ABC95214.1 </a>	TpmAC [uncultured bacterium]	<a href="#">43.5</a>	0.006

Top hit:

Display  Show  Send to

Range: from  to  Features:  CDD

l: [AAT78551](#). Reports thiopurine methyl...[gi:50542167]

[Features](#) [Sequence](#)

LOCUS AAT78551 65 aa linear BCT 08-FEB-2005  
DEFINITION thiopurine methyltransferase [Pseudomonas citronellolis].  
ACCESSION AAT78551  
VERSION AAT78551.1 GI:50542167  
DBSOURCE accession [AY616837.1](#)  
KEYWORDS .  
SOURCE Pseudomonas citronellolis  
ORGANISM [Pseudomonas citronellolis](#)  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (residues 1 to 65)  
AUTHORS Favre-Bonte,S., Ranjard,L., Colinson,C., Prigent-Combaret,C.,  
Nazaret,S. and Cournoyer,B.  
TITLE Freshwater selenium-methylating bacterial thiopurine  
methyltransferases: diversity and molecular phylogeny  
JOURNAL Environ. Microbiol. 7 (2), 153-164 (2005)  
PUBMED [15658983](#)  
REFERENCE 2 (residues 1 to 65)  
AUTHORS Favre-Bonte,S. and Cournoyer,B.  
TITLE Direct Submission  
JOURNAL Submitted (30-APR-2004) UMR CNRS 5557 Ecologie Microbienne,  
Universite Lyon 1, Villeurbanne 69622, France  
COMMENT Method: conceptual translation supplied by author.  
FEATURES Location/Qualifiers  
source 1..65  
/organism="Pseudomonas citronellolis"  
/strain="CFBP 5585"  
/db\_xref="taxon:[53408](#)"  
[Protein](#) 1..65  
/product="thiopurine methyltransferase"  
/name="TPMT"  
[Region](#) <1..>61  
/region\_name="TPMT"  
/note="Thiopurine S-methyltransferase (TPMT); pfam05724"  
/db\_xref="CDD:[45619](#)"  
[CDS](#) 1..65  
/gene="tpmG"  
/coded\_by="AY616837.1:<1..>197"  
/note="selenium methylation"  
/codon\_start=2  
/transl\_table=[11](#)  
ORIGIN  
1 ldlawlaaq levlgvelse kavsdffeeh dlhpeidqld gfrryrvagi tllqgdffac  
61 rqstw  
//

# Tax BLAST Report

## Index

- ◆ [Lineage Report](#)
- ◆ [Organism Report](#)
- ◆ [Taxonomy Report](#)
- ◆ [Help](#)

## Lineage Report

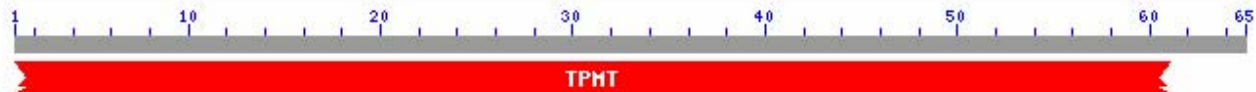
<u>root</u>			
cellular organisms			
Bacteria	[eubacteria]		
Proteobacteria	[proteobacteria]		
Gammaproteobacteria	[g-proteobacteria]		
Pseudomonadaceae	[g-proteobacteria]		
Pseudomonas	[g-proteobacteria]		
Pseudomonas aeruginosa group	[g-proteobacteria]		
Pseudomonas citronellois	136 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas citronellois]</a>
Pseudomonas aeruginosa PA01	121 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas aeruginosa PA01]</a>
Pseudomonas aeruginosa PACS2	118 1 hit [g-proteobacteria]		<a href="#">hypothetical protein PaerPA 01003444 [Pseudomonas aeruginosa]</a>
Pseudomonas aeruginosa UCEPP-PA14	115 1 hit [g-proteobacteria]		<a href="#">COG0500: SAM-dependent methyltransferases [Pseudomonas aeruginosa]</a>
Pseudomonas aeruginosa 2192	113 1 hit [g-proteobacteria]		<a href="#">COG0500: SAM-dependent methyltransferases [Pseudomonas aeruginosa]</a>
Pseudomonas aeruginosa PA7	104 1 hit [g-proteobacteria]		<a href="#">hypothetical protein PaerP 01004452 [Pseudomonas aeruginosa]</a>
Pseudomonas pseudoalcaligenes	67 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas pseudoalcaligenes]</a>
Pseudomonas asplenii	73 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas asplenii]</a>
Pseudomonas sp. Hsa.28	72 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas sp. Hsa.28]</a>
Pseudomonas fluorescens Pf-5	71 1 hit [g-proteobacteria]		<a href="#">thiopurine S-methyltransferase [Pseudomonas fluorescens Pf-5]</a>
Pseudomonas putida KT2440	68 1 hit [g-proteobacteria]		<a href="#">thiopurine S-methyltransferase [Pseudomonas putida KT2440]</a>
Pseudomonas putida F1	68 1 hit [g-proteobacteria]		<a href="#">thiopurine S-methyltransferase [Pseudomonas putida F1] &gt;gil</a>
Pseudomonas fluorescens Pf0-1	68 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Pseudomonas fluorescens Pf0-1]</a>
Pseudomonas fluorescens	66 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas fluorescens]</a>
Pseudomonas chlororaphis	64 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas chlororaphis]</a>
Pseudomonas fragi	64 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas fragi]</a>
Pseudomonas stutzeri	64 3 hits [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas stutzeri]</a>
Pseudomonas entomophila L48	62 1 hit [g-proteobacteria]		<a href="#">thiopurine S-methyltransferase [Pseudomonas entomophila L48]</a>
Pseudomonas viridiflava	61 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas viridiflava]</a>
Pseudomonas syringae pv. maculicola	58 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas syringae pv. maculicola]</a>
Pseudomonas syringae pv. tomato str. DC3000	58 1 hit [g-proteobacteria]		<a href="#">thiopurine S-methyltransferase [Pseudomonas syringae pv. tomato str. DC3000]</a>
Pseudomonas syringae pv. aptata	55 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas syringae pv. aptata]</a>
Pseudomonas syringae pv. japonica	54 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas syringae pv. japonica]</a>
Pseudomonas syringae pv. pisi	54 3 hits [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas syringae pv. pisi]</a>
Pseudomonas syringae pv. syringae B728a	54 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Pseudomonas syringae pv. syringae B728a]</a>
Pseudomonas syringae	53 1 hit [g-proteobacteria]		<a href="#">Chain A, Solution Structure Of Thiopurine Methyltransferase [Pseudomonas syringae]</a>
Pseudomonas syringae pv. dysoxylis	53 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas syringae pv. dysoxylis]</a>
Pseudomonas syringae pv. atrofaciens	51 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Pseudomonas syringae pv. atrofaciens]</a>
Azotobacter vinelandii AvOP	73 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Azotobacter vinelandii AvOP]</a>
Halomonas halophila	66 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Halomonas halophila]</a>
Thiomicrospira crunogena XCL-2	62 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Thiomicrospira crunogena XCL-2]</a>
Nahella chejuensis KCTC 2396	54 1 hit [g-proteobacteria]		<a href="#">SAM-dependent methyltransferase [Nahella chejuensis KCTC 2396]</a>
Marinomonas sp. MED121	53 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Marinomonas sp. MED121] &gt;gil8</a>
Nitrococcus mobilis Nb-231	53 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Nitrococcus mobilis Nb-231]</a>
Alteromonas macleodii 'Deep ecotype'	52 1 hit [g-proteobacteria]		<a href="#">thiopurine S-methyltransferase family protein [Alteromonas macleodii 'Deep ecotype']</a>
Shewanella sp. PV-4	51 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Shewanella sp. PV-4] &gt;gil78</a>
Shewanella amazonensis SB2B	51 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Shewanella amazonensis SB2B]</a>
Oceanospirillum sp. MED92	50 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Oceanospirillum sp. MED92]</a>
Xanthomonas campestris pv. vesicatoria str. 85-10	50 1 hit [g-proteobacteria]		<a href="#">thiopurine S-methyltransferase [Xanthomonas campestris pv. vesicatoria str. 85-10]</a>
Shewanella sp. MR-4	49 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Shewanella sp. MR-4] &gt;gil11</a>
Shewanella sp. MR-7	49 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Shewanella sp. MR-7] &gt;gil11</a>
Shewanella sp. ANA-3	48 1 hit [g-proteobacteria]		<a href="#">Thiopurine S-methyltransferase [Shewanella sp. ANA-3] &gt;gil7</a>
Xanthomonas axonopodis pv. citri str. 306	48 1 hit [g-proteobacteria]		<a href="#">thiopurine methyltransferase [Xanthomonas axonopodis pv. citri str. 306]</a>
Vibrio cholerae V51	47 1 hit [g-proteobacteria]		<a href="#">COG0500: SAM-dependent methyltransferases [Vibrio cholerae V51]</a>
Shewanella oneidensis MR-1	47 1 hit [g-proteobacteria]		<a href="#">thiopurine S-methyltransferase [Shewanella oneidensis MR-1]</a>
Vibrio cholerae 0395	46 1 hit [g-proteobacteria]		<a href="#">COG0500: SAM-dependent methyltransferases [Vibrio cholerae 0395]</a>




**Query sequence:** [[gi|50542167|gb|AAT78551.1|](#)]

thiopurine methyltransferase [*Pseudomonas citronellolis*]

Concise Result  
  Full Result  
  Show Search Information  
 [?](#)



## Descriptions


	Title	PssmId	Multi-Dom	E-value
	pfam05724, TPMT, Thiopurine S-methyltransferase (TPMT). This family consists of thiopu...	<a href="#">45619</a>	No	2e-16
	pfam05724, TPMT, Thiopurine S-methyltransferase (TPMT). This family consists of thiopurine S-methyltransferase proteins from both eukaryotes and prokaryotes. Thiopurine S-methyltransferase (TPMT) is a cytosolic enzyme that catalyses S-methylation of aromatic and heterocyclic sulfhydryl compounds, including anticancer and immunosuppressive thiopurines..			

**CD Length:** 218, **Pct. Aligned:** 27.981651, **Bit Score:** 78.0587, **E-value:** 2e-16

		10	20	30	40	50	60	
		....*	....*	....*	....*	....*	....*	....*
query	1	LDLAWLAAQGLEVLGVELSEKAVSDFFEEHDLHPEIDQLDGFRRYRVAGITLLQGDFFACR	61					
consensus	50	LDMVWLAEQGHFVVGVEISELAVEKFFAEANLSPHITELSFYRAGGITLLCGDFFTLTP	110					

[Search for similar domain architectures](#)

### CD Search Reference:

 Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", *Nucleic Acids Res.*32(W)327-331.



# Multiple sequence alignments for pfam05724.4



HOME SEARCH SITE MAP Entrez CDD Structure Protein Help

## pfam05724.4

### TPMT, with user query added

Thiopurine S-methyltransferase (TPMT). This family consists of thiopurine S-methyltransferase proteins from both eukaryotes and prokaryotes. Thiopurine S-methyltransferase (TPMT) is a cytosolic enzyme that catalyses S-methylation of aromatic and heterocyclic sulphhydryl compounds, including anticancer and immunosuppressive thiopurines.

[+] Links:

[+] Statistics:

Other Related Conserved Domains: [C060500](#) [C062226](#) [C062227](#) [C062230](#) [C062518](#) [pfam03848](#) [pfam05401](#)

Reformat Sequence Alignment Format: Compact Hypertext Row Display: All 12 rows Color Bits: 2.0 bits Type Selection: the most similar members

```


gi 12585364 19 .[50]. IEMKWFADRGHTVVGVEISEIGIRE.[2]. AEQNL.[3]. EEP    LTEIAGAKV.[15]. FD.[1]. P 134
query      1      LDLAWLAAQGLEVLGVELSEKAVSD.[2]. EEHDL    HPE    IDQLDGFRR.[13]. FA.[1]. R 61
gi 14583123 30 .[48]. HDVVAMASPERFVVGDLISESALEK.[2]. ETYGS    SPK    AKYFTFVKE.[ 2]. FT.[1]. R 127
gi 12585374 2 .[49]. EDLIWLANQHDSVQGVELSQIAVRS.[2]. AEHFY    TPT    VTRLNAQHE.[14]. FT.[1]. P 112
gi 9948918  1 .[49]. LDLAWLAAQGLEVLGVELSEKAVSD.[2]. EEHDL    HPE    IDQLDGFRR.[13]. FA.[1]. Q 110
gi 21112457 1 .[49]. LDLHWLAAQGHRLVGVESPLAVTQ.[2]. DDAGL    QPQ    RHTSRAGEH.[13]. FT.[1]. D 110
gi 6094506  1 .[49]. QDMSWLSGGQYHVVGAELESAEVER.[2]. TERGE    QPH    ITSQGFVKV.[13]. FA.[1]. T 110
gi 25136597 22 .[49]. LDMLWLAEQGHDIIGVELSPLAIEA.[2]. RENHL    PPS    KRRQGRFTL.[13]. FA.[1]. S 131
gi 13529134 24 .[50]. VEMKWFADRGHSVVGVEISELGIQE.[2]. TEQNL.[3]. EEP    ITEIPGTKV.[15]. FD.[1]. P 139
gi 17427472 13 .[47]. YEAGWLAERGWPVAAIDFAPS AVAS    AQA VL    GPH.[2]. VVELADFFR    FT    P 106
    
```

```

gi 12585364 135 .[106]. 240
query      62 .[ 4]. 65
gi 14583123 128 .[ 87]. 214
gi 12585374 113 .[105]. 217
gi 9948918  111 .[108]. 218
gi 21112457 111 .[108]. 218
gi 6094506  111 .[108]. 218
gi 25136597 132 .[108]. 239
gi 13529134 140 .[106]. 245
gi 17427472 107 .[101]. 207
    
```

Citing CDD: Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwatz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH (2005), "CDD: a Conserved Domain Database for protein classification.", *Nucleic Acids Res.* 33: D192-6

# Related Structure:

 **MMDB**  
**Structure Summary**

PubMed BLAST Structure Taxonomy OMIM **Help?** Cn3d

**Reference:** Scheuermann TH, Lolis E, Hodsdon ME [Tertiary structure of thiopurine methyltransferase from Pseudomonas syringae, a bacterial orthologue of a polymorphic, drug-metabolizing enzyme](#) *J. Mol. Biol.* v333, p.573-585


**Description:** Solution Structure Of Thiopurine Methyltransferase From Pseudomonas Syringae.

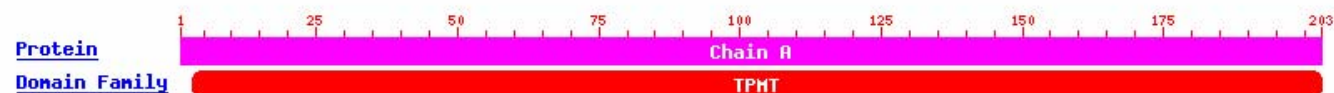
**Deposition:** 2003/6/4

**Taxonomy:** [Pseudomonas syringae](#)

**MMDB:** [25061](#) **PDB:** [1PJZ](#) **Structure Neighbors:** [VAST](#)

of

Molecular components in the MMDB structure are listed below. The icons indicate macromolecular chains, 3D domains, protein classifications and ligands. Please hold the mouse over each icon for more information on the component. 



**Please take a 5-min. survey before you leave!**

<http://www.mannlib.cornell.edu/>

- Reference and Instruction
  - Library Instruction
    - Workshops
      - Web survey



# BLAST: all

Sequences producing significant alignments:

		Score (Bits)	E Value
<a href="#">gi 1469268 emb CAA59282.1 </a>	firefly luciferase [Photinus pyralis]	<a href="#">1071</a>	0.0
<a href="#">gi 45384790 gb AAS59437.1 </a>	luciferase [Reporter vector pGSA13...]	<a href="#">1060</a>	0.0
<a href="#">gi 13160953 gb AAK13426.1 </a>	luciferase [Promoter probe vector pJB	<a href="#">1060</a>	0.0
<a href="#">gi 1469266 emb CAA59281.1 </a>	firefly luciferase [Photinus pyralis]	<a href="#">1058</a>	0.0
<a href="#">gi 3123921 gb AAC40214.1 </a>	firefly luciferase [Reporter vector p2	<a href="#">1058</a>	0.0
<a href="#">gi 209229 gb AAA72988.1 </a>	luciferase/kanamycin resistance protein	<a href="#">1054</a>	0.0
<a href="#">gi 37991672 dbj BAD00047.1 </a>	Fusion protein, Feo [Hepatitis C vir	<a href="#">1050</a>	0.0
<a href="#">gi 55535619 gb AAV52869.1 </a>	luciferase luc2 [Firefly luciferas...]	<a href="#">1050</a>	0.0
<a href="#">gi 69111702 gb AAZ03395.1 </a>	Aqp4-Luc fusion protein [Reporter vec	<a href="#">1050</a>	0.0
<a href="#">gi 69111664 gb AAZ03394.1 </a>	Aqp4-Luc fusion protein [Reporter vec	<a href="#">1050</a>	0.0
<a href="#">gi 55535625 gb AAV52873.1 </a>	luciferase luc2CP [Firefly lucifer...]	<a href="#">1050</a>	0.0
<a href="#">gi 55535622 gb AAV52871.1 </a>	luciferase luc2P [Firefly lucifera...]	<a href="#">1050</a>	0.0
<a href="#">gi 47420064 gb AAT27381.1 </a>	destabilized luciferase [Cloning v...]	<a href="#">1050</a>	0.0
<a href="#">gi 17530182 gb AAL40737.1 </a>	tissue factor/luciferase fusion prote	<a href="#">1050</a>	0.0
<a href="#">gi 17530179 gb AAL40735.1 </a>	protein serine kinase/luciferase f...	<a href="#">1050</a>	0.0
<a href="#">gi 7415877 dbj BAA93575.1 </a>	luciferase [synthetic construct]	<a href="#">1049</a>	0.0
<a href="#">gi 14290102 gb AAK59251.1 </a>	luciferase [Cloning vector pVLH/hs...]	<a href="#">1048</a>	0.0
<a href="#">gi 14009672 gb AAK51708.1 </a>	luciferase [Cloning vector pHLH/int(+	<a href="#">1048</a>	0.0
<a href="#">gi 42718124 gb AAS38485.1 </a>	luciferase [RNA interference vector p	<a href="#">1048</a>	0.0
<a href="#">gi 39653986 gb AAR29593.1 </a>	hlucCP+ reporter protein [Reporter ve	<a href="#">1048</a>	0.0
<a href="#">gi 39653983 gb AAR29591.1 </a>	hlucP+ reporter protein [Reporter vec	<a href="#">1048</a>	0.0
<a href="#">gi 14009669 gb AAK51706.1 </a>	luciferase [Cloning vector pVLH/int(+	<a href="#">1047</a>	0.0
<a href="#">gi 57634584 gb AAW52575.1 </a>	luciferase [Cloning vector p713-94...]	<a href="#">1046</a>	0.0
<a href="#">gi 52631875 gb AAU85360.1 </a>	luciferase [Lampyrus turkestanicus]	<a href="#">905</a>	0.0
<a href="#">gi 899315 emb CAA61668.1 </a>	photinus-luciferin 4-monooxygenase ...	<a href="#">902</a>	0.0
<a href="#">gi 54292811 gb AAV32457.1 </a>	luciferase [Cratomorphus distinctus]	<a href="#">895</a>	0.0
<a href="#">gi 38455132 gb AAR20792.1 </a>	luciferase [Pyrocoelia rufa]	<a href="#">878</a>	0.0
<a href="#">gi 12018178 gb AAG45439.1 </a>	luciferase [Pyrocoelia rufa]	<a href="#">874</a>	0.0
<a href="#">gi 38455134 gb AAR20793.1 </a>	luciferase [Pyrocoelia rufa]	<a href="#">869</a>	0.0
<a href="#">gi 695387 gb AAC37254.1 </a>	luciferase >gi 1584300 prf  2122369A lu	<a href="#">869</a>	0.0
<a href="#">gi 38455136 gb AAR20794.1 </a>	luciferase [Lampyrus noctiluca]	<a href="#">838</a>	0.0
<a href="#">gi 1669527 dbj BAA05006.1 </a>	luciferase [Photuris pennsylvanica]	<a href="#">753</a>	0.0
<a href="#">gi 1669525 dbj BAA05005.1 </a>	luciferase [Photuris pennsylvanica]	<a href="#">752</a>	0.0
<a href="#">gi 871401 emb CAA90072.1 </a>	luciferase [Luciola lateralis] >gi ...]	<a href="#">734</a>	0.0
<a href="#">gi 9527 emb CAA47358.1 </a>	luciferase [Luciola lateralis] >gi 26...]	<a href="#">734</a>	0.0
<a href="#">gi 1197516 emb CAA93444.1 </a>	luciferase [Luciola lateralis] >gi ...]	<a href="#">733</a>	0.0
<a href="#">gi 126500 sp P13129 LUCI LUCCR</a>	Luciferin 4-monooxygenase (Luc...]	<a href="#">733</a>	0.0
<a href="#">gi 30160347 gb AAO39673.2 </a>	luciferase type MJ1 [Luciola later...]	<a href="#">733</a>	0.0
<a href="#">gi 19880632 gb AAM00429.1 </a>	luciferase [Hotaria unmunšana]	<a href="#">723</a>	0.0
<a href="#">gi 24021175 gb AAN40978.1 </a>	luciferase [Hotaria papariensis] >...]	<a href="#">721</a>	0.0
<a href="#">gi 409317 gb AAB26932.1 </a>	luciferase [Luciola mingrellica] >gi ...]	<a href="#">721</a>	0.0
<a href="#">gi 24021171 gb AAN40976.1 </a>	luciferase [Hotaria tsushimana]	<a href="#">721</a>	0.0
<a href="#">gi 671718 gb AAC37253.1 </a>	luciferase >gi 1584301 prf  2122369B lu	<a href="#">720</a>	0.0
<a href="#">gi 24021177 gb AAN40979.1 </a>	luciferase [Hotaria tsushimana]	<a href="#">718</a>	0.0
<a href="#">gi 13094137 dbj BAB32737.1 </a>	luciferase [Cloning vector pPVLUC441	<a href="#">695</a>	0.0
<a href="#">gi 41688574 sp Q27757 LUCI PHOPE</a>	Luciferin 4-monooxygenase (L...]	<a href="#">638</a>	0.0
<a href="#">gi 4959885 gb AAD34542.1 </a>	luciferase [Phrixothrix vivianii]	<a href="#">587</a>	3e-166
<a href="#">gi 4959887 gb AAD34543.1 </a>	red-bioluminescence eliciting lucifera	<a href="#">523</a>	5e-147
<a href="#">gi 32455187 gb AAP83305.1 </a>	CBRLuc [Luciferase reporter vector...]	<a href="#">509</a>	1e-142
<a href="#">gi 32455193 gb AAP83309.1 </a>	CBG68luc [Luciferase reporter vect...]	<a href="#">505</a>	2e-141
<a href="#">gi 33333065 gb AAQ11697.1 </a>	luciferase [Pyrophorus plagiophthalam	<a href="#">504</a>	2e-141
<a href="#">gi 32455198 gb AAP83312.1 </a>	CBG99luc [Luciferase reporter vect...]	<a href="#">504</a>	3e-141

# BLAST - bacteria

Query ID: 1127935119-2281-123255581538.BLASTQ4

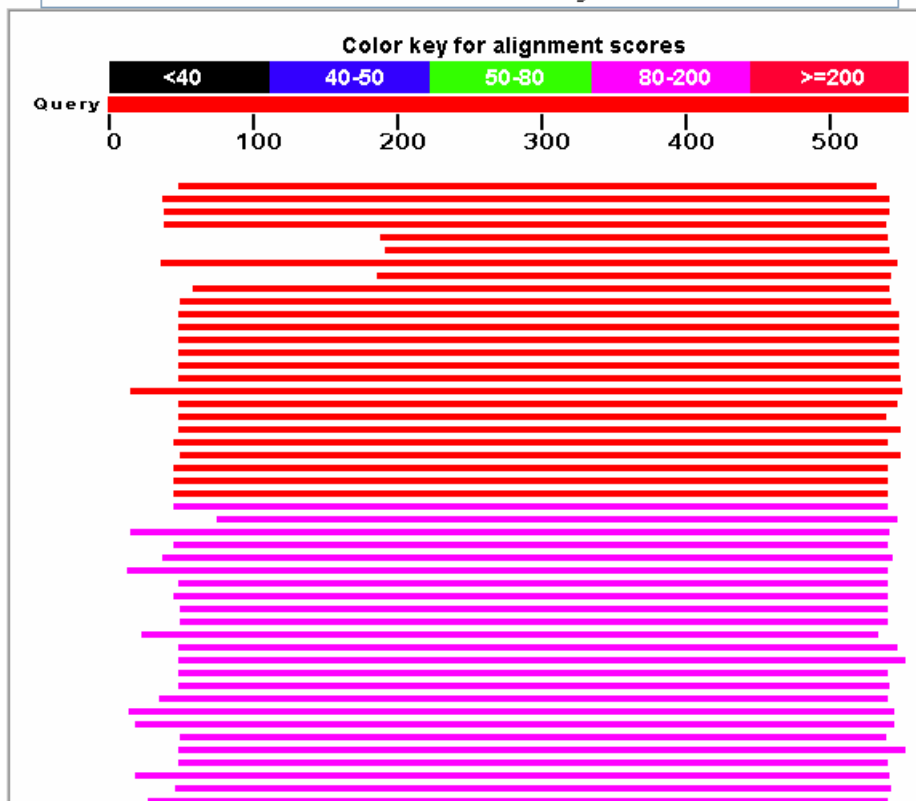
**Database:** All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples  
2,869,704 sequences; 984,312,347 total letters

If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)  
[Taxonomy reports](#)

**Query=**  
(550 letters)

## Distribution of 505 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments



# BLAST

- bacteria

Sequences producing significant alignments:

	Score (Bits)	E Value	
<a href="#">gi 23128059 ref ZP_00109915.1 </a>	<a href="#">266</a>	5e-70	
<a href="#">gi 8894733 emb CAB95894.1 </a>	<a href="#">261</a>	1e-68	G
<a href="#">gi 29607518 dbj BAC71576.1 </a>	<a href="#">252</a>	8e-66	G
<a href="#">gi 66870477 gb EAL97842.1 </a>	<a href="#">244</a>	2e-63	
<a href="#">gi 68263153 emb CAI36641.1 </a>	<a href="#">238</a>	2e-61	G
<a href="#">gi 56677440 gb AAV94106.1 </a>	<a href="#">234</a>	2e-60	G
<a href="#">gi 68262888 emb CAI36376.1 </a>	<a href="#">228</a>	2e-58	G
<a href="#">gi 68181452 ref ZP_00554435.1 </a>	<a href="#">219</a>	6e-56	
<a href="#">gi 56421317 ref YP_148635.1 </a>	<a href="#">210</a>	5e-53	G
<a href="#">gi 56419203 ref YP_146521.1 </a>	<a href="#">209</a>	8e-53	G
<a href="#">gi 49481470 ref YP_038577.1 </a>	<a href="#">209</a>	1e-52	G
<a href="#">gi 42739628 gb AAS43554.1 </a>	<a href="#">209</a>	1e-52	
<a href="#">gi 65321891 ref ZP_00394850.1 </a>	<a href="#">208</a>	1e-52	
<a href="#">gi 47530062 ref YP_021411.1 </a>	<a href="#">208</a>	1e-52	G
<a href="#">gi 51974447 gb AAU15997.1 </a>	<a href="#">208</a>	1e-52	G
<a href="#">gi 55981432 ref YP_144729.1 </a>	<a href="#">207</a>	4e-52	G
<a href="#">gi 51894152 ref YP_076843.1 </a>	<a href="#">206</a>	7e-52	G
<a href="#">gi 10175726 dbj BAB06823.1 </a>	<a href="#">206</a>	7e-52	G
<a href="#">gi 23099577 ref NP_693043.1 </a>	<a href="#">204</a>	3e-51	G
<a href="#">gi 46199401 ref YP_005068.1 </a>	<a href="#">203</a>	4e-51	G
<a href="#">gi 75759373 ref ZP_00739469.1 </a>	<a href="#">203</a>	6e-51	
<a href="#">gi 52004568 gb AAU24510.1 </a>	<a href="#">201</a>	2e-50	G
<a href="#">gi 30019243 ref NP_830874.1 </a>	<a href="#">201</a>	3e-50	G
<a href="#">gi 49477085 ref YP_035342.1 </a>	<a href="#">200</a>	5e-50	G
<a href="#">gi 42736187 gb AAS40123.1 </a>	<a href="#">200</a>	5e-50	G
<a href="#">gi 65318473 ref ZP_00391432.1 </a>	<a href="#">199</a>	1e-49	
<a href="#">gi 30022603 ref NP_834234.1 </a>	<a href="#">198</a>	1e-49	G
<a href="#">gi 52002767 gb AAU22709.1 </a>	<a href="#">197</a>	2e-49	G
<a href="#">gi 51977690 gb AAU19240.1 </a>	<a href="#">197</a>	2e-49	G
<a href="#">gi 68178669 ref ZP_00551779.1 </a>	<a href="#">197</a>	3e-49	
<a href="#">gi 68170294 gb EAM98182.1 </a>	<a href="#">197</a>	4e-49	
<a href="#">gi 51891757 ref YP_074448.1 </a>	<a href="#">197</a>	4e-49	G
<a href="#">gi 47567565 ref ZP_00238276.1 </a>	<a href="#">197</a>	4e-49	
<a href="#">gi 67153689 ref ZP_00415434.1 </a>	<a href="#">196</a>	5e-49	
<a href="#">gi 23098124 ref NP_691590.1 </a>	<a href="#">196</a>	5e-49	G
<a href="#">gi 68270143 ref ZP_00576391.1 </a>	<a href="#">196</a>	7e-49	
<a href="#">gi 56421225 ref YP_148543.1 </a>	<a href="#">196</a>	9e-49	G
<a href="#">gi 71547316 ref ZP_00667969.1 </a>	<a href="#">195</a>	2e-48	
<a href="#">gi 10175725 dbj BAB06822.1 </a>	<a href="#">195</a>	2e-48	G
<a href="#">gi 73537875 ref YP_298242.1 </a>	<a href="#">194</a>	3e-48	G
<a href="#">gi 16125218 ref NP_419782.1 </a>	<a href="#">193</a>	6e-48	G
<a href="#">gi 39982970 gb AAR34429.1 </a>	<a href="#">192</a>	1e-47	G
<a href="#">gi 6449061 gb AAF08801.1 </a>	<a href="#">192</a>	1e-47	
<a href="#">gi 68228889 ref ZP_00568089.1 </a>	<a href="#">192</a>	1e-47	

# BLAST- bacteria

## Tax BLAST Report

### Index

- ◆ [Lineage Report](#)
- ◆ [Organism Report](#)
- ◆ [Taxonomy Report](#)
- ◆ [Help](#)

### Lineage Report

<a href="#">Bacteria</a>	[eubacteria]				
- <a href="#">Cyanobacteria</a>	[cyanobacteria]				
.. <a href="#">Nostocaceae</a>	[cyanobacteria]				
... <a href="#">Nostoc punctiforme</a>	PCC 73102	266	2 hits	[cyanobacteria]	<a href="#">C060318: Acyl-CoA synthetases (AMP-forming)/AMP-acid ligase</a>
... <a href="#">Anabaena variabilis</a>	ATCC 29413	187	1 hit	[cyanobacteria]	<a href="#">AMP-dependent synthetase and ligase [Anabaena variabilis AT probable long-chain fatty-acid-CoA ligase [Gloeobacter viol 4-coumarate:CoA ligase [Streptomyces coelicolor A3(2)] &gt;gil putative 4-coumarate:CoA ligase [Streptomyces avermitilis M</a>
... <a href="#">Gloeobacter violaceus</a>	PCC 7421	190	3 hits	[cyanobacteria]	<a href="#">AMP-dependent synthetase and ligase [Arthrobacter sp. FB24] acyl-CoA synthetase [Corynebacterium jeikeium K411] &gt;gil685</a>
... <a href="#">Streptomyces coelicolor</a>	A3(2)	261	3 hits	[high GC Gram+]	<a href="#">4-coumarate:CoA ligase [Silicibacter pomeroyi DSS-3] &gt;gil56</a>
... <a href="#">Streptomyces avermitilis</a>	MA-4680	252	6 hits	[high GC Gram+]	<a href="#">AMP-dependent synthetase and ligase [Jannaschia sp. CCS1] &gt; long-chain fatty-acid-CoA ligase [Geobacillus kaustophilus</a>
... <a href="#">Arthrobacter</a>	sp. FB24	244	4 hits	[high GC Gram+]	<a href="#">long-chain-fatty-acid--CoA ligase (long-chain acyl-CoA synt</a>
... <a href="#">Corynebacterium jeikeium</a>	K411	238	4 hits	[high GC Gram+]	<a href="#">long-chain-fatty-acid--CoA ligase [Bacillus cereus ATCC 109</a>
... <a href="#">Silicibacter pomeroyi</a>	DSS-3	234	2 hits	[a-proteobacteria]	<a href="#">C060318: Acyl-CoA synthetases (AMP-forming)/AMP-acid ligase</a>
... <a href="#">Jannaschia</a>	sp. CCS1	219	2 hits	[a-proteobacteria]	<a href="#">long-chain-fatty-acid--CoA ligase (long-chain acyl-CoA synt</a>
... <a href="#">Geobacillus kaustophilus</a>	HTA426	210	11 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase (long-chain acyl-CoA synt</a>
... <a href="#">Bacillus thuringiensis</a>	serovar konkukian str. 97-27	209	7 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Bacillus cereus ATCC 109</a>
... <a href="#">Bacillus cereus</a>	ATCC 10987	209	7 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase (long-chain acyl-CoA synt</a>
... <a href="#">Bacillus anthracis</a>	str. A2012	208	2 hits	[eubacteria]	<a href="#">C060318: Acyl-CoA synthetases (AMP-forming)/AMP-acid ligase</a>
... <a href="#">Bacillus anthracis</a>	str. Ames	208	8 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Bacillus anthracis str.</a>
... <a href="#">Bacillus cereus</a>	E33L	208	8 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase (long-chain acyl-CoA synt</a>
... <a href="#">Thermus thermophilus</a>	HB8	207	2 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Thermus thermophilus HB8</a>
... <a href="#">Symbiobacterium thermophilum</a>	IAM 14863	206	4 hits	[high GC Gram+]	<a href="#">long-chain-fatty-acid--CoA ligase [Symbiobacterium thermophil</a>
... <a href="#">Bacillus halodurans</a>	C-125	206	5 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Bacillus halodurans C-125</a>
... <a href="#">Oceanobacillus iheyensis</a>	HTE831	204	7 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Oceanobacillus iheyensis</a>
... <a href="#">Thermus thermophilus</a>	HB27	203	4 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Thermus thermophilus HB27</a>
... <a href="#">Bacillus thuringiensis</a>	serovar israelensis ATCC 35646	203	4 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Bacillus thuringiensis s</a>
... <a href="#">Bacillus licheniformis</a>	ATCC 14580	201	7 hits	[eubacteria]	<a href="#">long chain acyl-CoA synthetase [Bacillus licheniformis ATCC</a>
... <a href="#">Bacillus cereus</a>	ATCC 14579	201	6 hits	[eubacteria]	<a href="#">Long-chain-fatty-acid--CoA ligase [Bacillus cereus ATCC 145</a>
... <a href="#">Desulfuromonas acetoxidans</a>	DSM 584	197	1 hit	[d-proteobacteria]	<a href="#">AMP-dependent synthetase and ligase [IMP dehydrogenase/GMP r</a>
... <a href="#">Desulfitobacterium hafniense</a>	DCE-2	197	5 hits	[eubacteria]	<a href="#">AMP-dependent synthetase and ligase [Desulfitobacterium haf</a>
... <a href="#">Bacillus cereus</a>	G9241	197	7 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Bacillus cereus G9241] &gt;</a>
... <a href="#">Azotobacter vinelandii</a>	AvOP	196	2 hits	[g-proteobacteria]	<a href="#">AMP-dependent synthetase and ligase [Azotobacter vinelandii</a>
... <a href="#">Moorella thermoacetica</a>	ATCC 39073	196	3 hits	[eubacteria]	<a href="#">AMP-dependent synthetase and ligase [Moorella thermoacetica</a>
... <a href="#">Syntrophobacter fumaroxidans</a>	MP0B	195	4 hits	[d-proteobacteria]	<a href="#">AMP-dependent synthetase and ligase [Syntrophobacter fumaro</a>
... <a href="#">Ralstonia eutropha</a>	JMP134	194	18 hits	[b-proteobacteria]	<a href="#">AMP-dependent synthetase and ligase [Ralstonia eutropha JMP</a>
... <a href="#">Caulobacter crescentus</a>	CB15	193	3 hits	[a-proteobacteria]	<a href="#">long-chain-fatty-acid--CoA ligase, putative [Caulobacter cr</a>
... <a href="#">Geobacter sulfurreducens</a>	PCA	192	1 hit	[d-proteobacteria]	<a href="#">long-chain-fatty-acid--CoA ligase, putative [Geobacter sulf</a>
... <a href="#">Bacillus subtilis</a>		192	6 hits	[eubacteria]	<a href="#">YngI [Bacillus subtilis]</a>
... <a href="#">Frankia</a>	sp. EANLpeC	192	5 hits	[high GC Gram+]	<a href="#">AMP-dependent synthetase and ligase [Frankia sp. EANLpeC] &gt;</a>
... <a href="#">Bacillus clausii</a>	KSM-K16	191	3 hits	[eubacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Bacillus clausii KSM-K16</a>
... <a href="#">Vibrio</a>	sp. Ex25	191	2 hits	[g-proteobacteria]	<a href="#">C060318: Acyl-CoA synthetases (AMP-forming)/AMP-acid ligase</a>
... <a href="#">Bdellovibrio bacteriovorus</a>	HD100	190	2 hits	[d-proteobacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Bdellovibrio bacteriovoru</a>
... <a href="#">Exiguobacterium</a>	sp. 255-15	189	2 hits	[eubacteria]	<a href="#">AMP-dependent synthetase and ligase [Exiguobacterium sp. 25</a>
... <a href="#">Escherichia coli</a>	O157:H7 EDL933	188	1 hit	[enterobacteria]	<a href="#">long-chain-fatty-acid--CoA ligase [Escherichia coli O157:H7</a>
... <a href="#">Ralstonia metallidurans</a>	CH34	187	17 hits	[b-proteobacteria]	<a href="#">AMP-dependent synthetase and ligase [Ralstonia metalliduran</a>
... <a href="#">Vibrio parahaemolyticus</a>	RIMD 2210633	187	2 hits	[g-proteobacteria]	<a href="#">putative long-chain-fatty-acid--CoA ligase [Vibrio parahaemo</a>
... <a href="#">Rubrobacter xylanophilus</a>	DSM 9941	187	5 hits	[high GC Gram+]	<a href="#">AMP-dependent synthetase and ligase [Rubrobacter xylanophil</a>
... <a href="#">Shigella flexneri</a>	2a str. 301	187	1 hit	[enterobacteria]	<a href="#">acyl-CoA synthetase, long-chain-fatty-acid--CoA ligase [Shi</a>
... <a href="#">Brevibacterium linens</a>	BL2	186	4 hits	[high GC Gram+]	<a href="#">C060318: Acyl-CoA synthetases (AMP-forming)/AMP-acid ligase</a>
... <a href="#">Escherichia coli</a>		186	3 hits	[enterobacteria]	<a href="#">acyl-CoA synthetase, long-chain-fatty-acid--CoA ligase [Esc</a>
... <a href="#">Solibacter usitatus</a>	Ellin6076	186	3 hits	[eubacteria]	<a href="#">AMP-dependent synthetase and ligase [Solibacter usitatus El</a>
... <a href="#">Shigella sonnei</a>	Ss046	186	1 hit	[enterobacteria]	<a href="#">acyl-CoA synthetase [Shigella sonnei Ss046] &gt;gil73855360 gb</a>
... <a href="#">Escherichia coli</a>	CFT073	186	1 hit	[enterobacteria]	<a href="#">Long-chain-fatty-acid--CoA ligase [Escherichia coli CFT073]</a>
... <a href="#">Escherichia coli</a>	E24377A	186	1 hit	[enterobacteria]	<a href="#">C060318: Acyl-CoA synthetases (AMP-forming)/AMP-acid ligase</a>
... <a href="#">Shigella boydii</a>	BS512	186	1 hit	[enterobacteria]	<a href="#">C060318: Acyl-CoA synthetases (AMP-forming)/AMP-acid ligase</a>