IGS Annotation Engine and Manatee

Michelle Gwinn Giglio

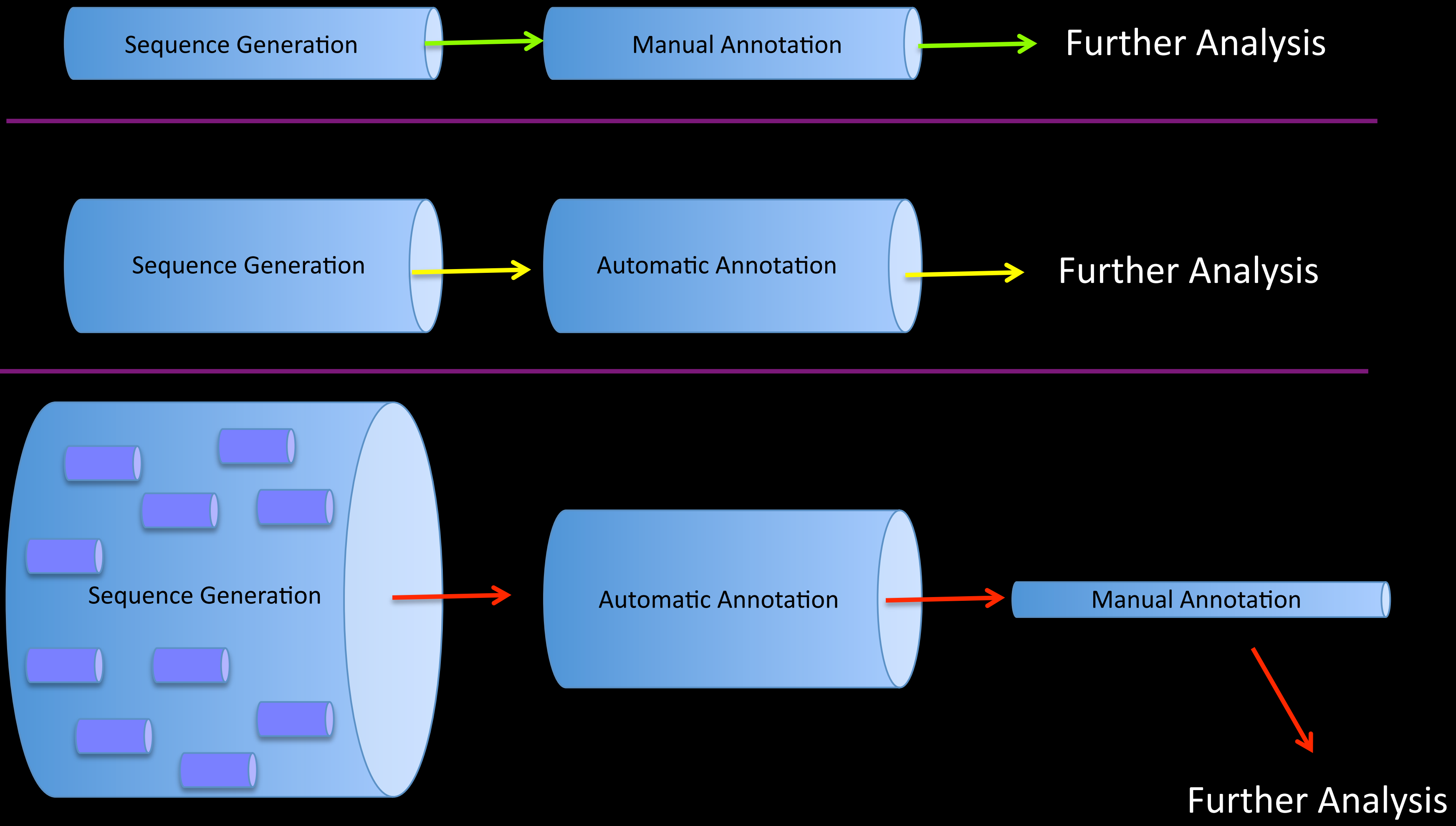
Pathway Tools Workshop

October 2010

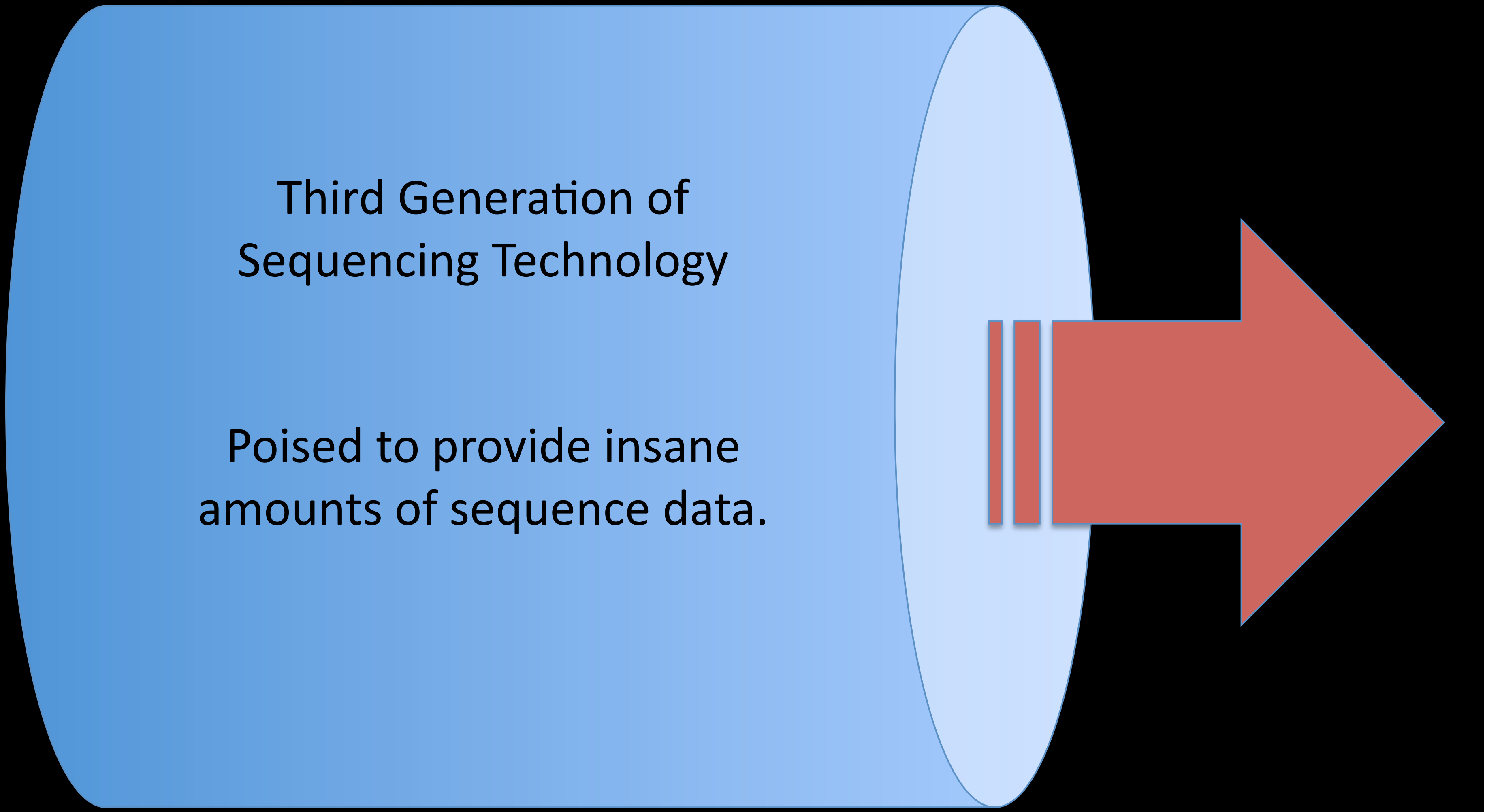
IGS Annotation Engine

- A free service to anyone with a prokaryotic sequence they wish to annotate that provides:
  - Automated output of the IGS prokaryotic annotation pipeline
  - The Manatee curation tool
- Can be used with complete or draft genomes

The need for services like the AE



More is on the way!!!

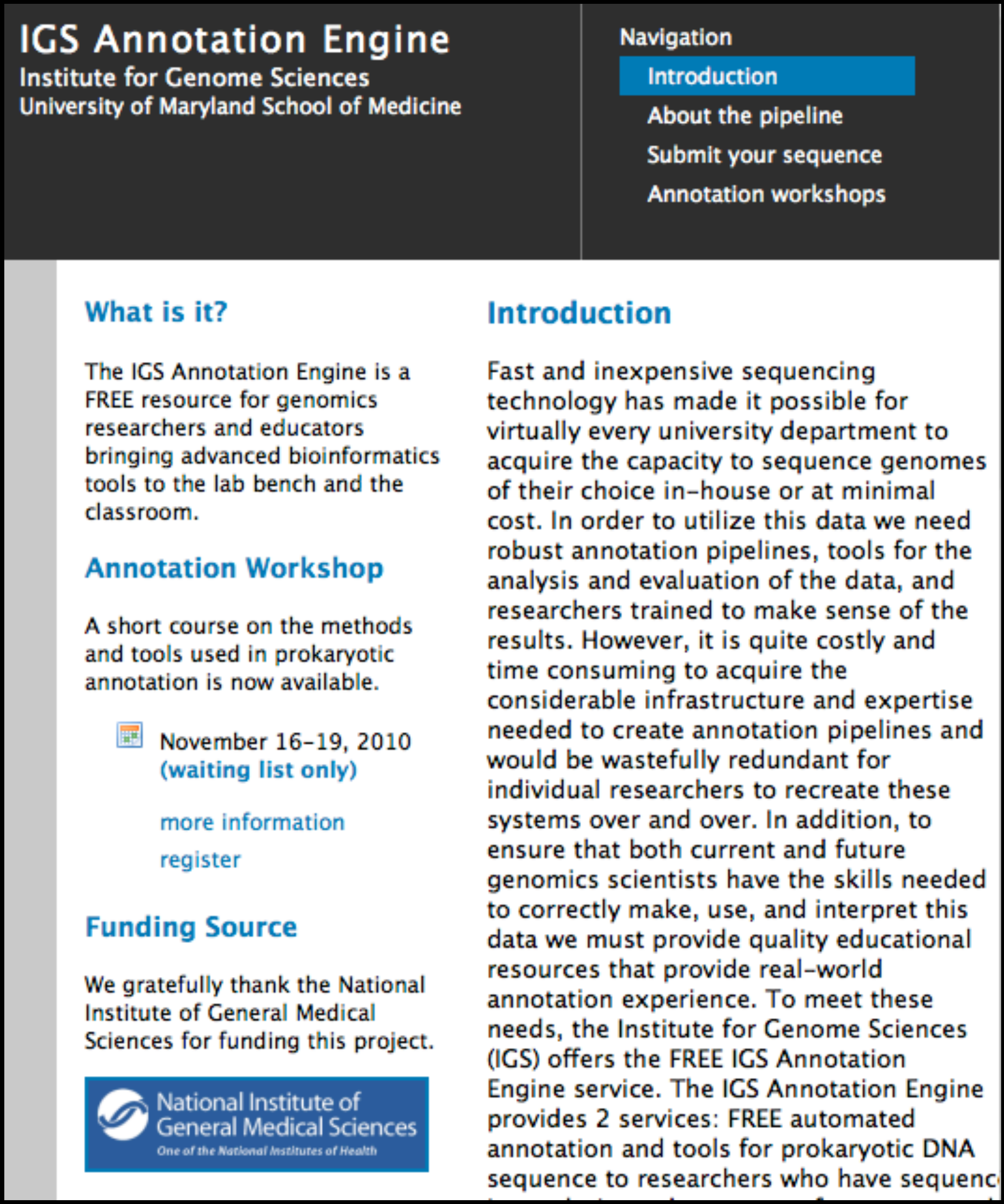


Third Generation of Sequencing Technology

Poised to provide insane amounts of sequence data.

Annotation Engine web page

http://ae.igs.umaryland.edu



IGS Annotation Engine

Institute for Genome Sciences
University of Maryland School of Medicine

What is it?

The IGS Annotation Engine is a FREE resource for genomics researchers and educators bringing advanced bioinformatics tools to the lab bench and the classroom.

Annotation Workshop

A short course on the methods and tools used in prokaryotic annotation is now available.

November 16–19, 2010 **(waiting list only)**

more information

register

Funding Source

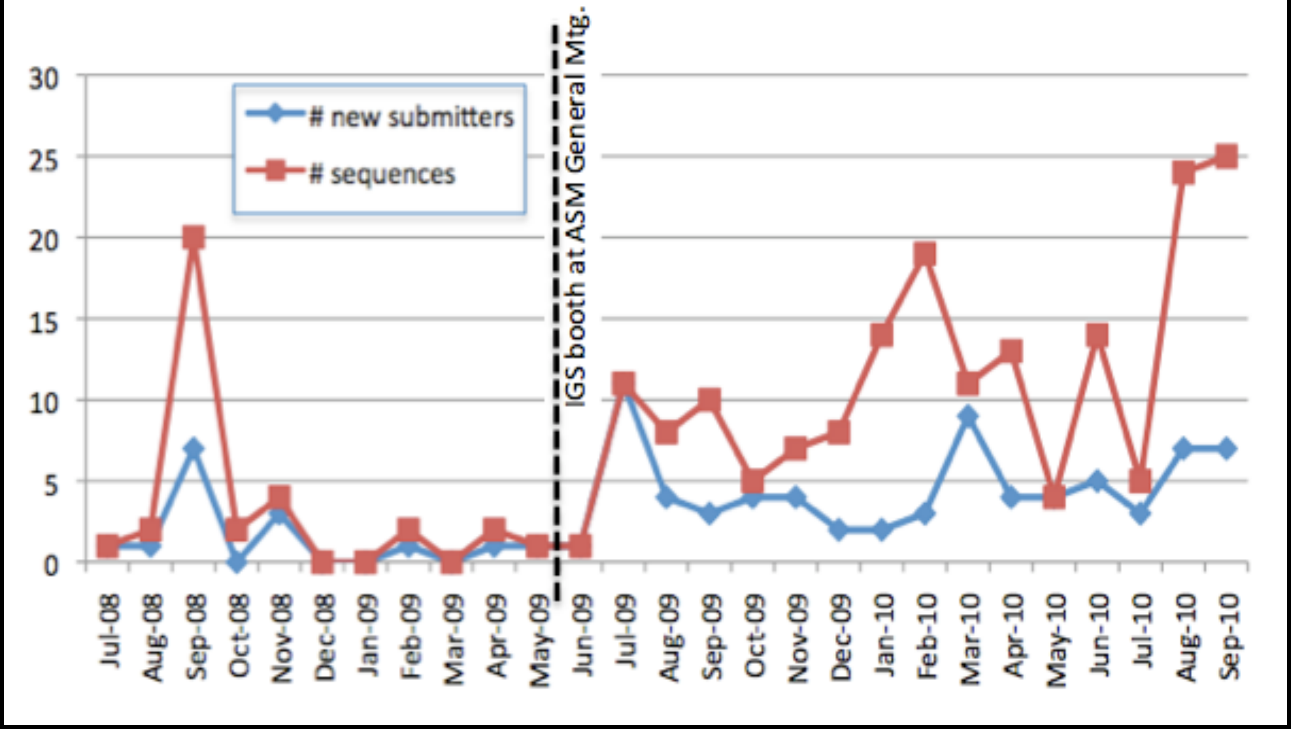
We gratefully thank the National Institute of General Medical Sciences for funding this project.

National Institute of General Medical Sciences
One of the National Institutes of Health

Introduction

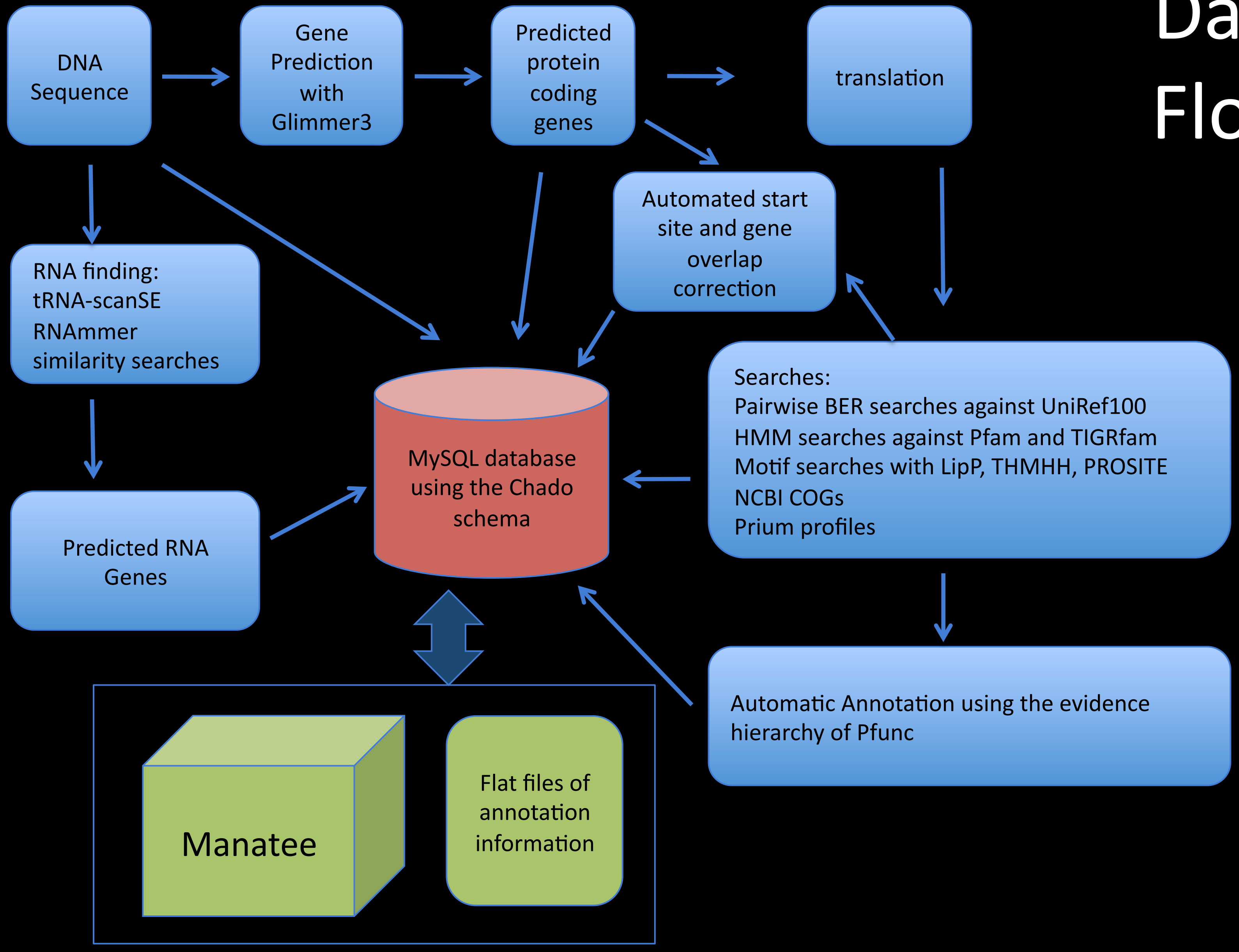
Fast and inexpensive sequencing technology has made it possible for virtually every university department to acquire the capacity to sequence genomes of their choice in-house or at minimal cost. In order to utilize this data we need robust annotation pipelines, tools for the analysis and evaluation of the data, and researchers trained to make sense of the results. However, it is quite costly and time consuming to acquire the considerable infrastructure and expertise needed to create annotation pipelines and would be wastefully redundant for individual researchers to recreate these systems over and over. In addition, to ensure that both current and future genomics scientists have the skills needed to correctly make, use, and interpret this data we must provide quality educational resources that provide real-world annotation experience. To meet these needs, the Institute for Genome Sciences (IGS) offers the FREE IGS Annotation Engine service. The IGS Annotation Engine provides 2 services: FREE automated annotation and tools for prokaryotic DNA sequence to researchers who have sequenc

IGS Annotation Engine Growth



- Current stats (from the two years of the project at IGS)
  - Submitters: 90
    - From all over the United States and 17 other countries
  - Users: >> 90
  - Genomes/sequences: >225

Data Flow



DNA Sequence

Gene Prediction with Glimmer3

Predicted protein coding genes

translation

Automated start site and gene overlap correction

RNA finding:
tRNA-scanSE
RNAmmer
similarity searches

MySQL database using the Chado schema

Searches:
Pairwise BER searches against UniRef100
HMM searches against Pfam and TIGRfam
Motif searches with LipP, THMHH, PROSITE
NCBI COGs
Prium profiles

Predicted RNA Genes

Automatic Annotation using the evidence hierarchy of Pfunc

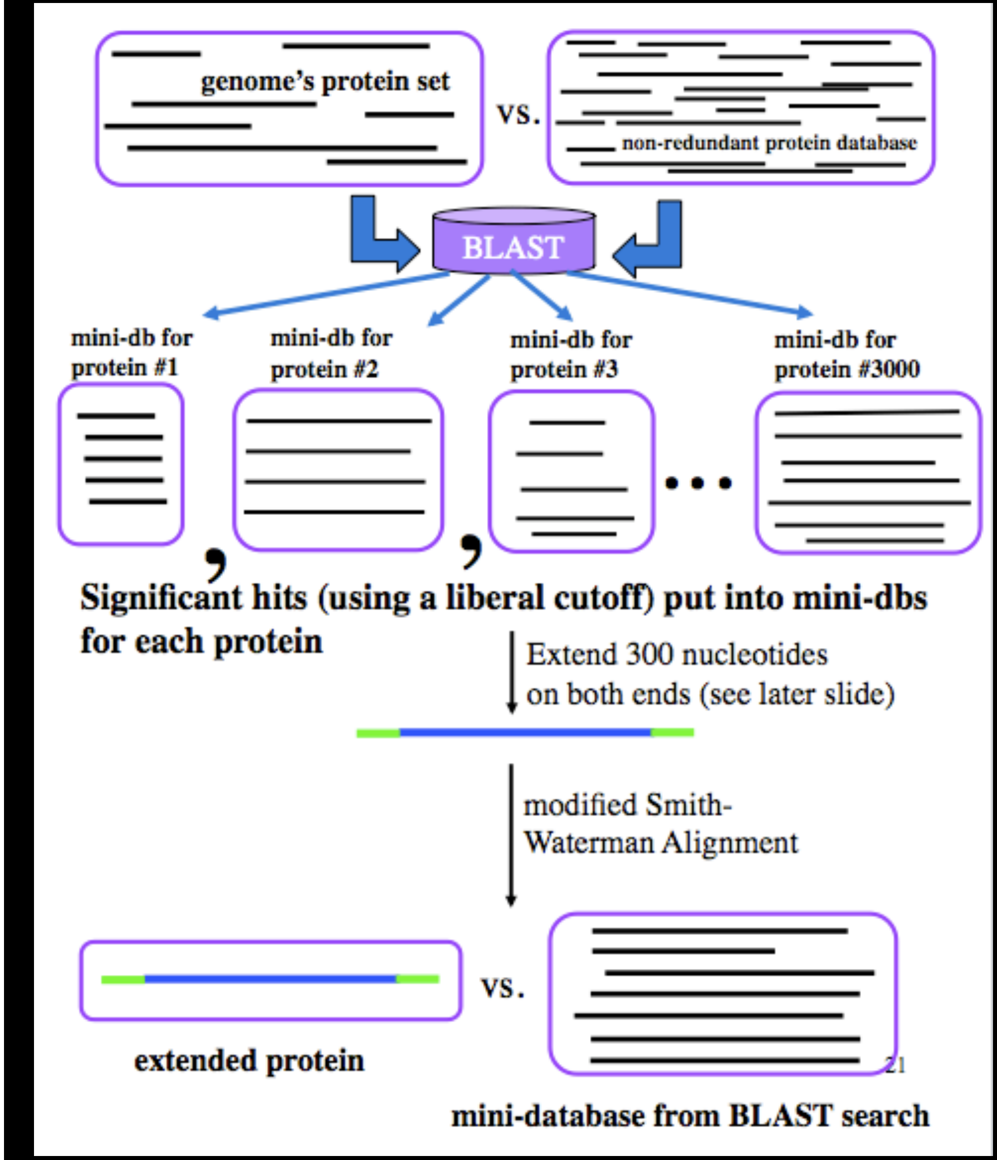
Manatee

Flat files of annotation information

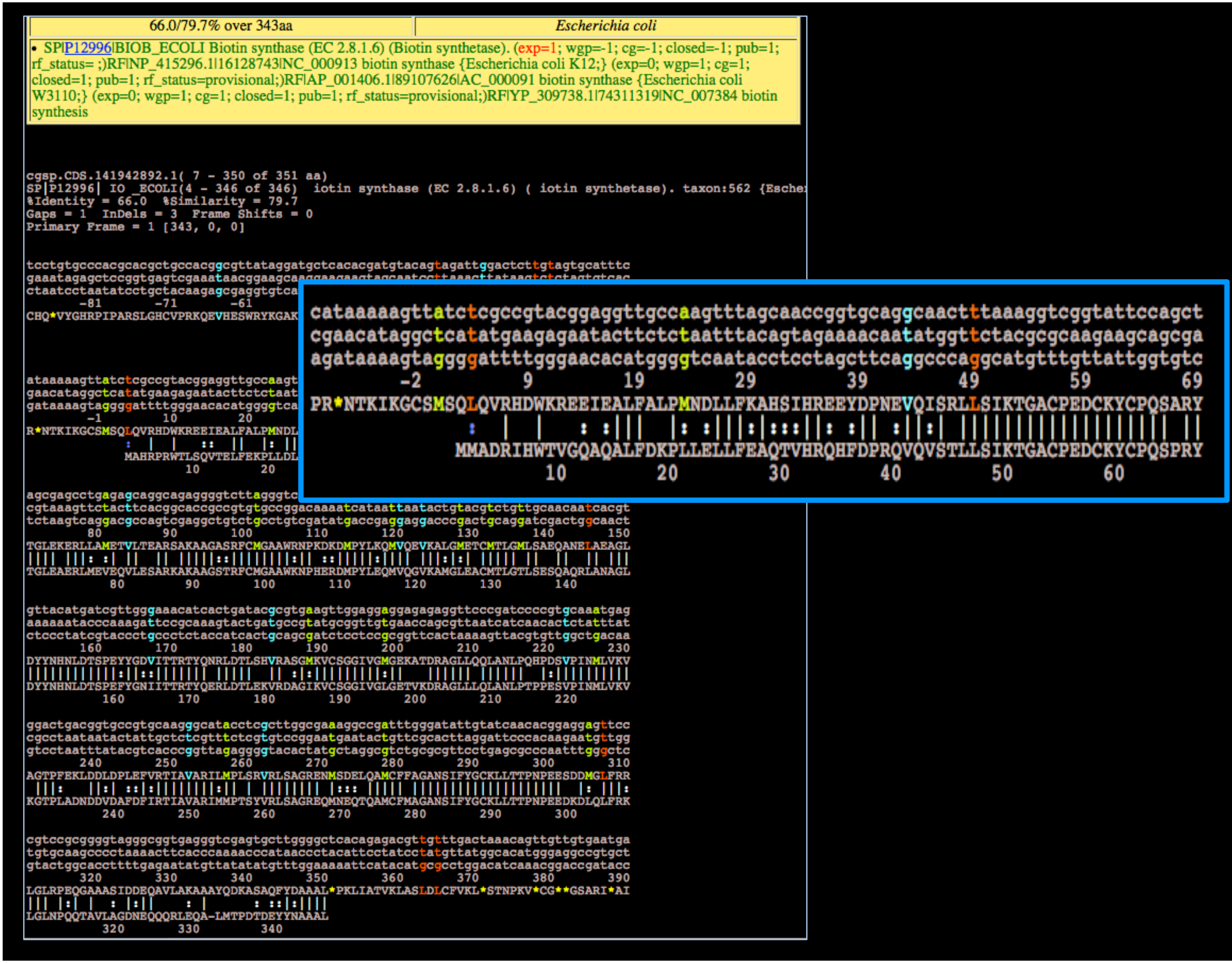
Sequence-based searches

- Pairwise protein alignments
- HMM searches
- Motif searches
  - PROSITE
  - TMHMM
  - SignalP
  - LipoP
- COGs
- Priam profiles

Blast-Extend-Repraze (BER)


genome's protein set
vs.
non-redundant protein database
BLAST
mini-db for protein #1
mini-db for protein #2
mini-db for protein #3
mini-db for protein #3000
Significant hits (using a liberal cutoff) put into mini-dbs for each protein
Extend 300 nucleotides on both ends (see later slide)
modified Smith-Waterman Alignment
extended protein
vs.
mini-database from BLAST search
21

- a pairwise alignment tool
- initial BLAST with liberal cutoff for each protein in the genome
- modified Smith-Waterman alignment generated between search protein and each BLAST result
- result is a file containing one pairwise alignment for each match protein from the BLAST
- view alignments in our Manatee annotation tool
- we do the 2-step process because BLAST is fast and Smith-Waterman is slow, so it saves cpu time to only do the Smith-Waterman alignments on things that have any hope of matching

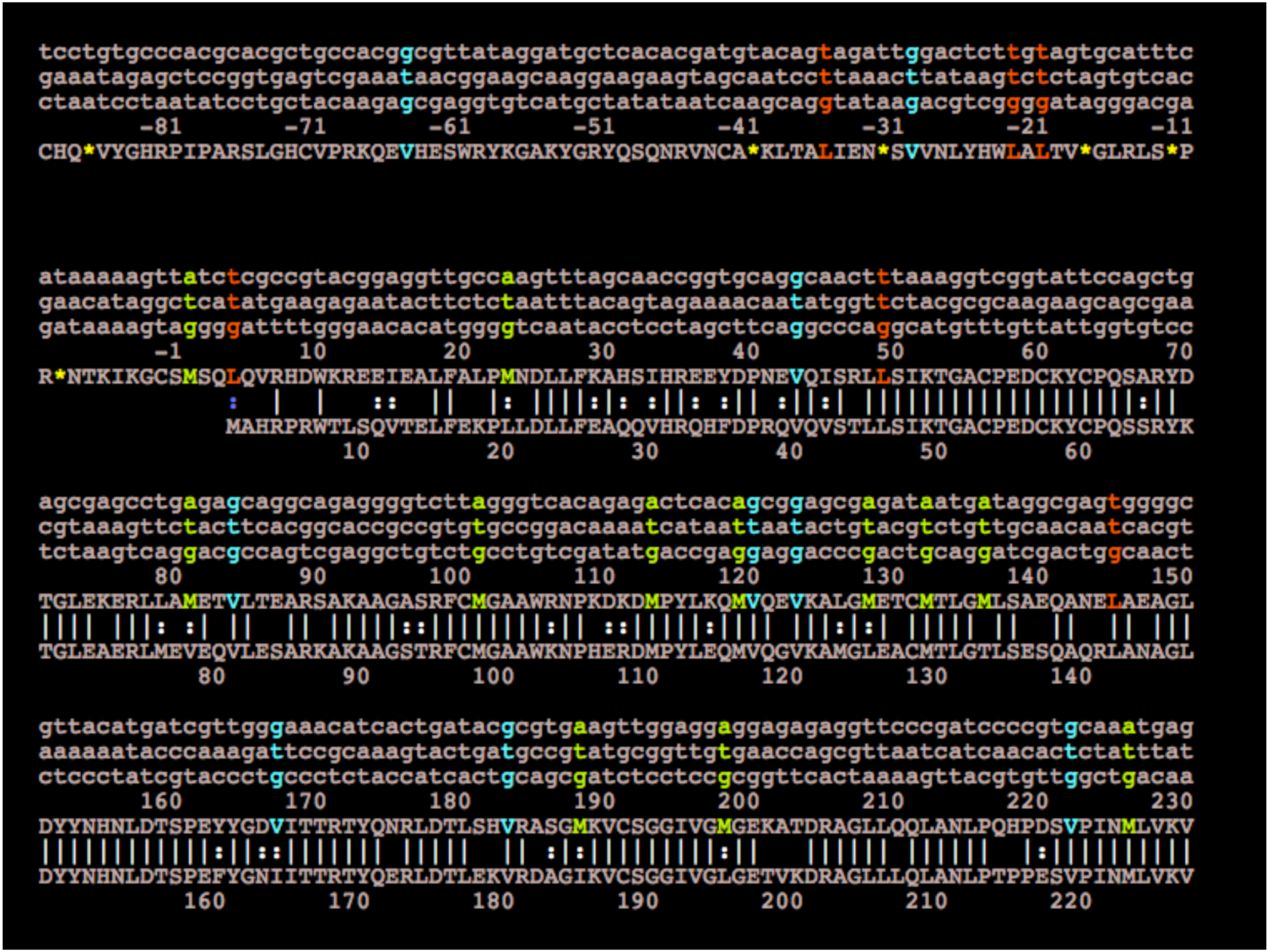


66.0/79.7% over 343aa	*Escherichia coli*

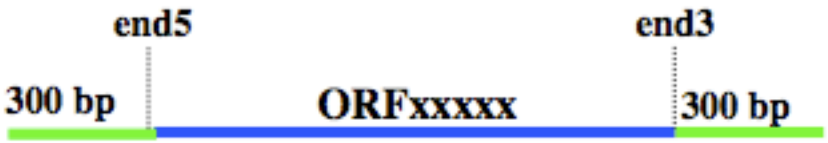
- SP|P12996|BIOB_ECOLI Biotin synthase (EC 2.8.1.6) (Biotin synthetase). (exp=1; wgp=-1; cg=-1; closed=-1; pub=1; rf_status= ;)RF|NP_415296.1|16128743|NC_000913 biotin synthase {Escherichia coli K12;} (exp=0; wgp=1; cg=1; closed=1; pub=1; rf_status=provisional;)RF|AP_001406.1|89107626|AC_000091 biotin synthase {Escherichia coli W3110;} (exp=0; wgp=1; cg=1; closed=1; pub=1; rf_status=provisional;)RF|YP_309738.1|74311319|NC_007384 biotin synthesis

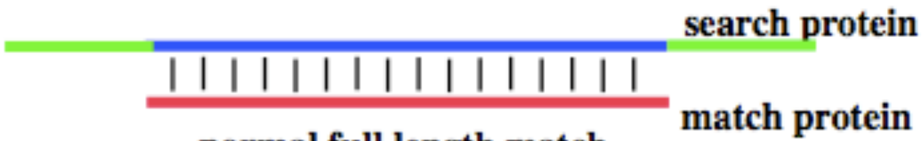
```
cgsp.CDS.141942892.1( 7 - 350 of 351 aa)
SP|P12996| IO _ECOLI(4 - 346 of 346)  iotin synthase (EC 2.8.1.6) ( iotin synthetase). taxon:562 {Esche
%Identity = 66.0  %Similarity = 79.7
Gaps = 1  InDels = 3  Frame Shifts = 0
Primary Frame = 1 [343, 0, 0]


tcctgtgcccacgcacgctgccacggcgttataggatgctcacacgatgtacagtagattggactcttgtagtgcatttc
gaaatagagctccggtgagtcgaaataacggaagcaaggaagaagtagcaatccttaaacttataagtctctagtgtcac
ctaatcctaatatcctgctacaagagcgaggtgtca
       -81       -71       -61
CHQ*VYGHRPIPARSLGHCVPRKQEVHESWRYKGAK



ataaaaagttatctcgccgtacggaggttgccaagt
gaacataggctcatatgaagagaatacttctctaat
gataaaagtaggggatttgggaacacatggggtca
        -1        10        20
R*NTKIKGCSMSQLQVRHDWKREEIEALFALPMNDL
             :  |  |   ::  ||  |: ||
             MAHRPRWTLSQVTELFEKPLLDL
                     10        20

agcgagcctgagagcaggcagaggggtcttagggtc
cgtaaagttctacttcacggcaccgccgtgtgccggacaaaatcataattaatactgtacgtctgttgcaacaatcacgt
tctaagtcaggacgccagtcgaggctgtctgcctgtcgatatgaccgaggaggacccgactgcaggatcgactggcaact
        80        90       100       110       120       130       140       150
TGLEKERLLAMETVLTEARSAKAAGASRFCMGAAWRNPKDKDMPYLKQMVQEVKALGMETCMTLGMLSAEQANELAEAGL
|||| |||: :| ||  || |||||::||||||||:|| ::|||||:|||| |||:|:| ||||| ||  ||  || |||
TGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAWKNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGL
           80        90       100       110       120       130       140

gttacatgatcgttgggaaacatcactgatacgcgtgaagttggaggaggagagaggttcccgatccccgtgcaaatgag
aaaaaatacccaaagattccgcaaagtactgatgccgtatgcggttgtgaaccagcgttaatcatcaacactctatttat
ctccctatcgtaccctgccctctaccatcactgcagcgatctcctccgcggttcactaaaagttacgtgttggctgacaa
       160       170       180       190       200       210       220       230
DYYNHNLDTSPEYYGDVITTRTYQNRLDTLSHVRASGMKVCSGGIVGMGEKATDRAGLLQQLANLPQHPDSVPINMLVKV
||||||||||||:||::||||||| |||||  || :|:|||||||||:||   |||||| ||||||  |:||||||||||
DYYNHNLDTSPEFYGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPESVPINMLVKV
          160       170       180       190       200       210       220

ggactgacggtgccgtgcaagggcatacctcgcttggcgaaaggccgatttgggatattgtatcaacacggaggagttcc
cgcctaataatactattgctctcgtttctcgtgtccggaatgaatactgttcgcacttaggattcccacaagaatgttgg
gtcctaatttatacgtcacccggttagagggtacactatgctaggcgtctgcgcgttcctgagcgcccaatttgggctc
       240       250       260       270       280       290       300       310
AGTPFEKLDDLDPLEFVRTIAVARILMPLSRVRLSAGRENMSDELQAMCFFAGANSIFYGCKLLTTPNPEESDDMGLFRR
 |||:   ||:| ::|:||||||||:|| | |||||||| |::: ||||| ||||||||||||||||||||  |: |||:
KGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMCFMAGANSIFYGCKLLTTPNPEEDKDLQLFRK
          240       250       260       270       280       290       300

cgtccgcggggtagggcggtgagggtcgagtgcttggggctcacagagacgttgtttgactaaacagttgttgtgaatga
tgtgcaagcccctaaaacttcacccaaaacccataaccctacattcctatcctatgttatggcacatgggaggccgtgct
gtactggcaccttttgagaatatgttatatatgtttggaaaaattcatacatgcgcctggacatcaaacggaccgatacc
       320       330       340       350       360       370       380       390
LGLRPEQGAAASIDDEQAVLAKAAAYQDKASAQFYDAAAL*PKLIATVKLASLDLCFVKL*STNPKV*CG**GSARI*AI
||| |:| |  : |:||    : |      : ::|:||||
LGLNPQQTAVLAGDNEQQQRLEQA-LMTPDTDEYYNAAAL
          320       330       340
```

```
cataaaaagttatctcgccgtacggaggttgccaagtttagcaaccggtgcaggcaactttaaaggtcggtattccagct
cgaacataggctcatatgaagagaatacttctctaatttacagtagaaaacaatatggttctacgcgcaagaagcagcga
agataaaagtaggggatttgggaacacatggggtcaatacctcctagcttcaggcccaggcatgtttgttattggtgtc
        -2         9        19        29        39        49        59        69
PR*NTKIKGCSMSQLQVRHDWKREEIEALFALPMNDLLFKAHSIHREEYDPNEVQISRLLSIKTGACPEDCKYCPQSARY
              :  |  |   : :|||  |: :|||:|:::||: :|| :||:| ||||||||||||||||||| ||
             MMADRIHWTVGQAQALFDKPLLELLFEAQTVHRQHFDPRQVQVSTLLSIKTGACPEDCKYCPQSPRY
                     10        20        30        40        50        60
```

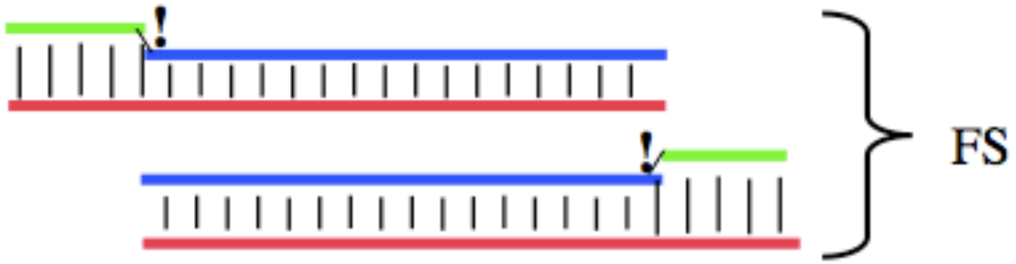


```
tcctgtgcccacgcacgctgccacggcgttataggatgctcacacgatgtacagtagattggactcttgtagtgcatttc
gaaatagagctccggtgagtcgaaataacggaagcaaggaagaagtagcaatccttaaacttataagtctctagtgtcac
ctaatcctaatatcctgctacaagagcgaggtgtcatgctatataatcaagcaggtataagacgtcggggatagggacga
       -81       -71       -61       -51       -41       -31       -21       -11
CHQ*VYGHRPIPARSLGHCVPRKQEVHESWRYKGAKYGRYQSQNRVNCA*KLTALIEN*SVVNLYHWLALTV*GLRLS*P


ataaaaagttatctcgccgtacggaggttgccaagtttagcaaccggtgcaggcaactttaaaggtcggtattccagctg
gaacataggctcatatgaagagaatacttctctaatttacagtagaaaacaatatggttctacgcgcaagaagcagcgaa
gataaaagtaggggattttgggaacacatggggtcaatacctcctagcttcaggcccaggcatgtttgttattggtgtcc
        -1        10        20        30        40        50        60        70
R*NTKIKGCSMSQLQVRHDWKREEIEALFALPMNDLLFKAHSIHREEYDPNEVQISRLLSIKTGACPEDCKYCPQSARYD
             :  |  |   ::  ||  |: ||||:|: :||: :|| :||:| |||||||||||||||||||:||
             MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVSTLLSIKTGACPEDCKYCPQSSRYK
                     10        20        30        40        50        60

agcgagcctgagagcaggcagaggggtcttagggtcacagagactcacagcggagcgagataatgataggcgagtggggc
cgtaaagttctacttcacggcaccgccgtgtgccggacaaaatcataattaatactgtacgtctgttgcaacaatcacgt
tctaagtcaggacgccagtcgaggctgtctgcctgtcgatatgaccgaggaggacccgactgcaggatcgactggcaact
        80        90       100       110       120       130       140       150
TGLEKERLLAMETVLTEARSAKAAGASRFCMGAAWRNPKDKDMPYLKQMVQEVKALGMETCMTLGMLSAEQANELAEAGL
|||| |||: :| ||  || |||||::||||||||:|| ::|||||:|||| |||:|:| ||||| ||  ||  || |||
TGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAWKNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGL
           80        90       100       110       120       130       140

gttacatgatcgttgggaaacatcactgatacgcgtgaagttggaggaggagagaggttcccgatccccgtgcaaatgag
aaaaaatacccaaagattccgcaaagtactgatgccgtatgcggttgtgaaccagcgttaatcatcaacactctatttat
ctccctatcgtaccctgccctctaccatcactgcagcgatctcctccgcggttcactaaaagttacgtgttggctgacaa
       160       170       180       190       200       210       220       230
DYYNHNLDTSPEYYGDVITTRTYQNRLDTLSHVRASGMKVCSGGIVGMGEKATDRAGLLQQLANLPQHPDSVPINMLVKV
||||||||||||:||::||||||| |||||  || :|:|||||||||:||   |||||| ||||||  |:||||||||||
DYYNHNLDTSPEFYGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPESVPINMLVKV
          160       170       180       190       200       210       220
```

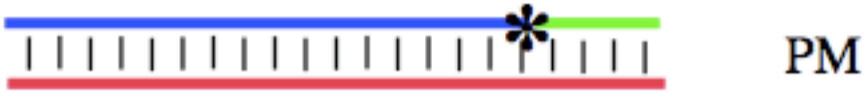




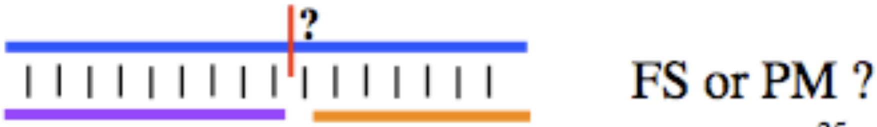
normal full length match



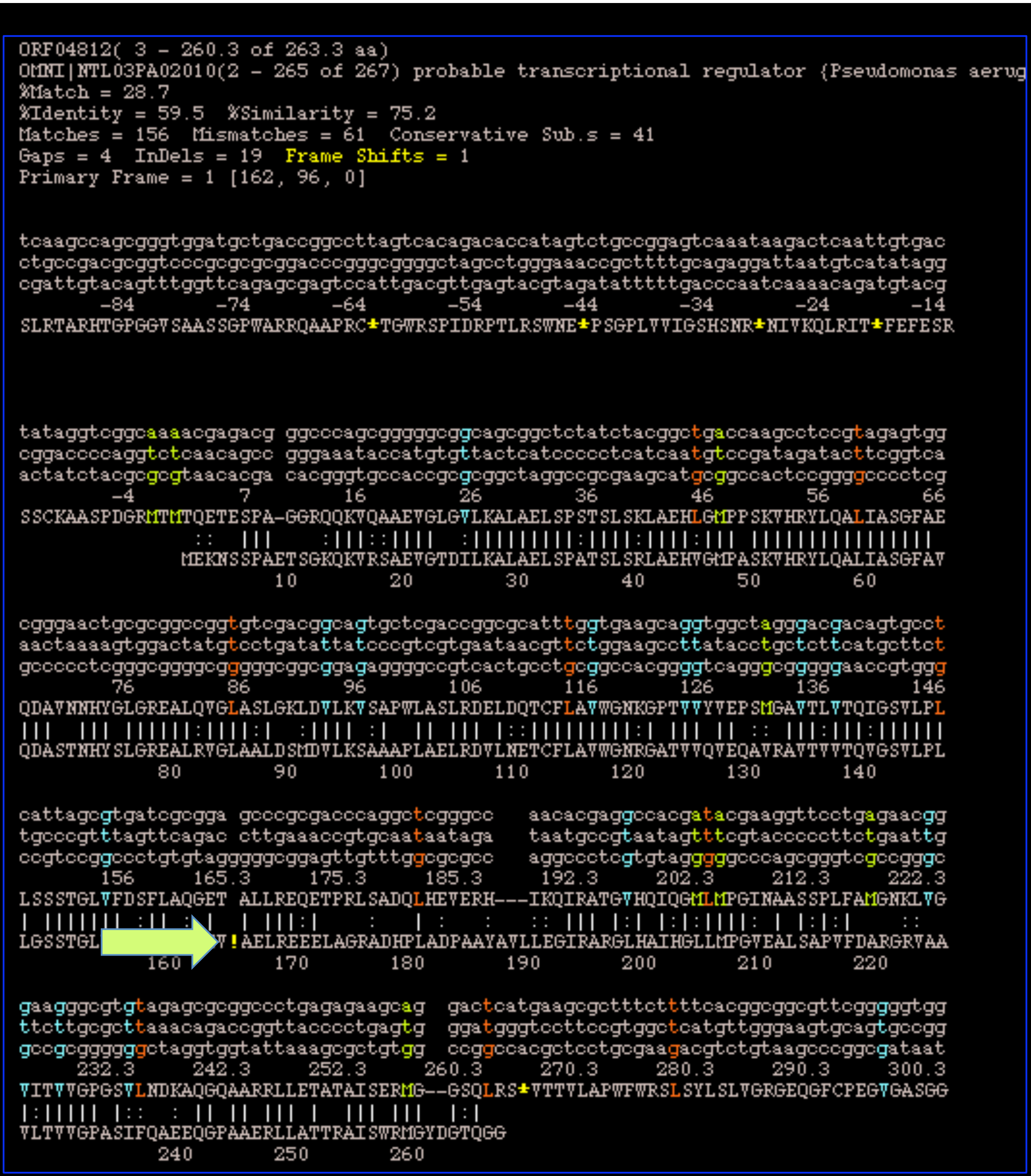
similarity extending through a frameshift upstream or downstream into extensions



similarity extending in the same frame through a stop codon



two functionally unrelated genes from other species matching one query protein could indicate incorrectly fused ORFs



```
ORF04812( 3 - 260.3 of 263.3 aa)
OMNI|NTL03PA02010(2 - 265 of 267) probable transcriptional regulator {Pseudomonas aerug
%Match = 28.7
%Identity = 59.5  %Similarity = 75.2
Matches = 156  Mismatches = 61  Conservative Sub.s = 41
Gaps = 4  InDels = 19  Frame Shifts = 1
Primary Frame = 1 [162, 96, 0]


tcaagccagcgggtggatgctgaccggccttagtcacagacaccatagtctgccggagtcaaataagactcaattgtgac
ctgccgacgcggtcccgcgcgcggacccgggcggggctagcctgggaaaccgcttttgcagaggattaatgtcatatagg
cgattgtacagtttggttcagagcgagtccattgacgttgagtacgtagatatttttgacccaatcaaaacagatgtacg
       -84       -74       -64       -54       -44       -34       -24       -14
SLRTARHTGPGGVSAASSGPWARRQAAPRC*TGWRSPIDRPTLRSWNE*PSGPLVVIGSHSNR*NIVKQLRIT*FEFESR



tataggtcggcaaaacgagacg ggcccagcgggggcggcagcggctctatctacggctgaccaagcctccgtagagtgg
cggaccccaggtctcaacagcc gggaaataccatgtgttactcatcccctcatcaatgtccgatagatacttcggtca
actatctacgcgcgtaacacga cacgggtgccaccgcgcggctaggccgcgaagcatgcggccactccggggcccctcg
        -4         7        16        26        36        46        56        66
SSCKAASPDGRMTMTQETESPA-GGRQQKVQAAEVGLGVLKALAELSPSTSLSKLAEHLGMPPSKVHRYLQALIASGFAE
               ::  |||    :|||::||||  :|||||||||:||||:||||:||| ||||||||||||||||
              MEKNSSPAETSGKQKVRSAEVGTDILKALAELSPATSLSRLAEHVGMPASKVHRYLQALIASGFAV
                      10        20        30        40        50        60

cgggaactgcgcggccggtgtcgacggcagtgctcgaccggcgcatttggtgaagcaggtggctagggacgacagtgcct
aactaaaagtggactatgtcctgatattatcccgtcgtgaataacgttctggaagccttatacctgctcttcatgcttct
gccccctcgggcggggcggggcggcggagaggggccgtcactgcctgcggccacgggtcagggcggggaaccgtggg
        76        86        96       106       116       126       136       146
QDAVNNHYGLGREALQVGLASLGKLDVLKVSAPWLASLRDELDQTCFLAVWGNKGPTVVYVEPSMGAVTLVTQIGSVLPL
|||  ||| ||||||:||||:|  :|||| :|  || ||| |::|||||||||:| ||| ||  ::|||:|||:||||||
QDASTNHYSLGREALRVGLAALDSMDVLKSAAAPLAELRDVLNETCFLAVWGNRGATVVQVEQAVRAVTVVTQVGSVLPL
            80        90       100       110       120       130       140

cattagcgtgatcgcgga gcccgcgacccaggctcgggcc   aacacgaggccacgatacgaaggttcctgagaacgg
tgcccgtttagttcagac cttgaaaccgtgcaataataga   taatgccgtaatagtttcgtaccccttctgaattg
ccgtccggccctgtgtagggggcggagttgtttggcgcgcc   aggccctcgtgtagggggcccagcgggtcgccgggc
       156     165.3     175.3     185.3     192.3     202.3     212.3     222.3
LSSSTGLVFDSFLAQGET ALLREQETPRLSADQLHEVERH---IKQIRATGVHQIQGMLMPGINAASSPLFAMGNKLVG
| ||||||| :|| : |  | |||:|    :   | :   :   :: ||| |:| |:|:||||: | |:|:|    ::
LGSSTGL          V!AELREEELAGRADHPLADPAAYAVLLEGIRARGLHAIHGLLMPGVEALSAPVFDARGRVAA
           160       170       180       190       200       210       220

gaagggcgtgtagagcgcggccctgagagaagcag  gactcatgaagcgctttcttttcacggcggcgttcgggggtgg
ttcttgcgcttaaacagaccggttacccctgagtg  ggatgggtccttccgtggctcatgttgggaagtgcagtgccgg
gccgcggggggctaggtggtattaaagcgctgtgg  ccggccacgctcctgcgaagacgtctgtaagcccggcgataat
     232.3     242.3     252.3     260.3     270.3     280.3     290.3     300.3
VITVVGPGSVLNDKAQGQAARRLLETATAISERMG--GSQLRS*VTTVLAPWFWRSLSYLSLVGRGEQGFCPEGVGASGG
|:||||| |::  : || || ||| |  ||| |||  |:|
VLTVVGPASIFQAEEQGPAAERLLATTRAISWRMGYDGTQGG
           240       250       260
```

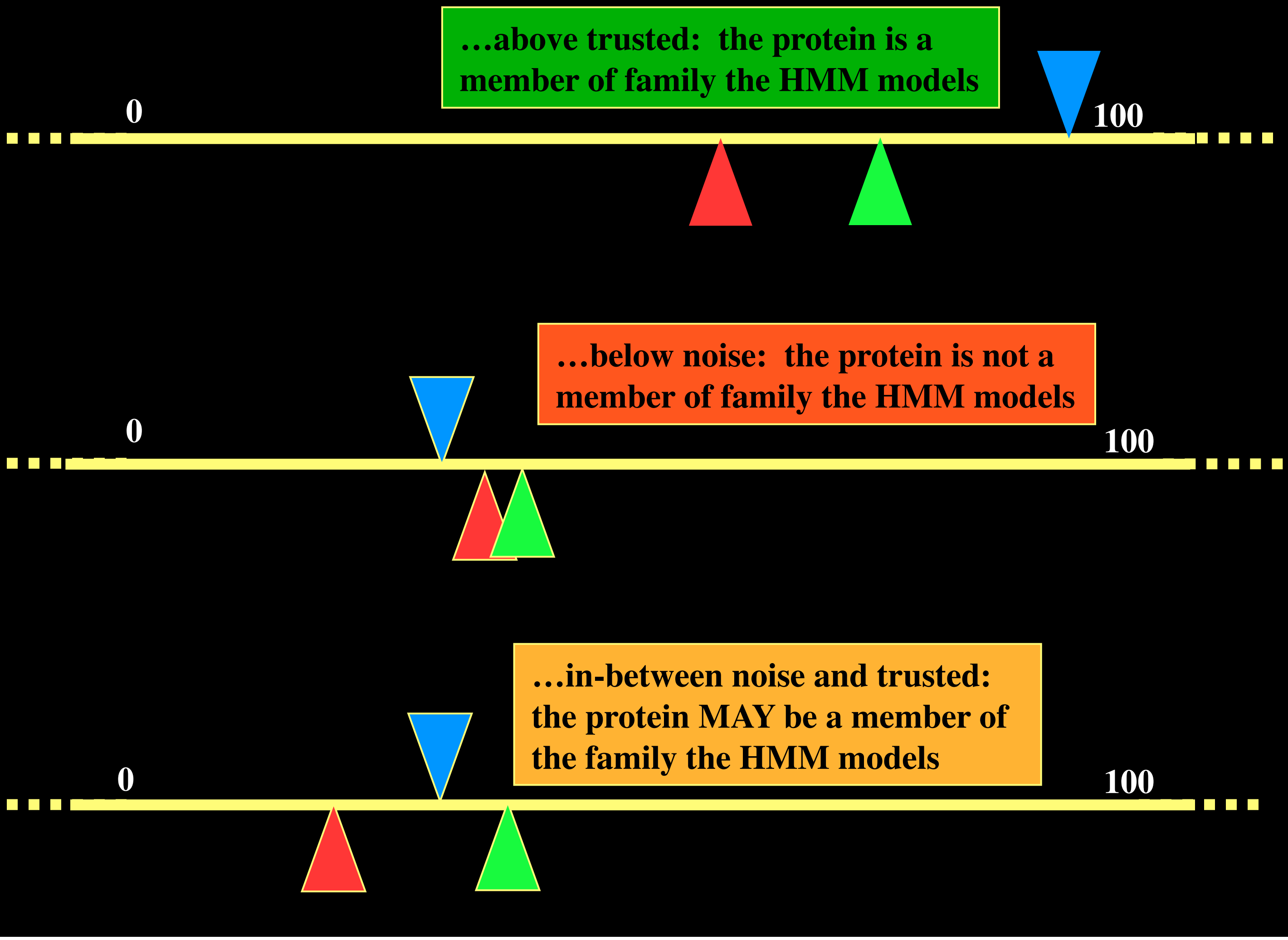
HMMs

- Our Hidden Markov Model database consists of TIGRFAMs and Pfam
- statistical model of the patterns of amino acids in a multiple alignment of proteins (called the “seed) which share sequence and functional similarity
- Each TIGRFAM HMM is assigned to a category which describes the type of relationship the proteins in the model have to each other
  - equivalog
  - superfamily
  - subfamily
  - domain
- one can search proteins against HMMs, they receive a score indicating how well they match the model
- by comparing this score to the cutoff scores assigned to each model, one can determine whether or not the search protein is a member of the group defined by the HMM
  - “trusted cutoff’ - proteins scoring above this score are considered a member of the group defined by the HMM
  - “noise cutoff” - proteins scoring below this score are considered NOT to be a member of the group defined by the HMM
  - for proteins scoring between trusted and noise, the HMM evidence is not sufficient to determine whether the protein is a member of the functional group or not

Annotation is attached to HMMs

- TIGR00433
  - category: equivalog
  - name: biotin synthase
  - EC: 2.8.1.6
  - gene symbol: bioB
  - GO terms: GO:0004076 biotin synthase activity; GO:0009102 biotin biosynthesis
- PF04055
  - category: domain
  - name: radical SAM domain protein
  - EC: not applicable
  - gene symbol: not applicable
  - GO terms: GO:0003824 catalytic activity; GO:0008152 metabolism

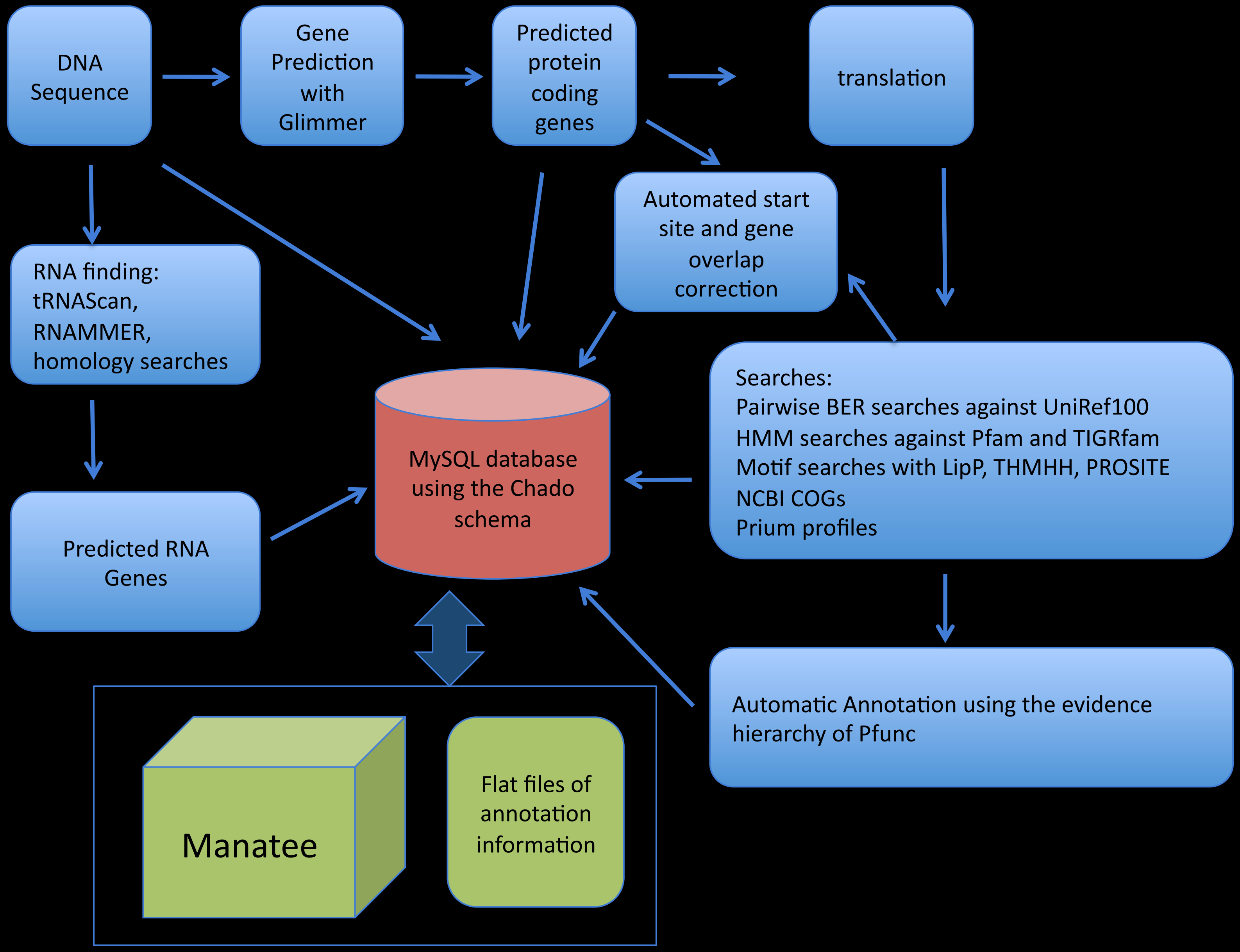
Evaluating HMM scores



…above trusted: the protein is a member of family the HMM models

…below noise: the protein is not a member of family the HMM models

…in-between noise and trusted: the protein MAY be a member of the family the HMM models



DNA Sequence

Gene Prediction with Glimmer

Predicted protein coding genes

translation

RNA finding: tRNAScan, RNAMMER, homology searches

Automated start site and gene overlap correction

Predicted RNA Genes

MySQL database using the Chado schema

Searches:
Pairwise BER searches against UniRef100
HMM searches against Pfam and TIGRfam
Motif searches with LipP, THMHH, PROSITE
NCBI COGs
Prium profiles

Automatic Annotation using the evidence hierarchy of Pfunc

Manatee

Flat files of annotation information

The Pitfalls of Transitive Annotation

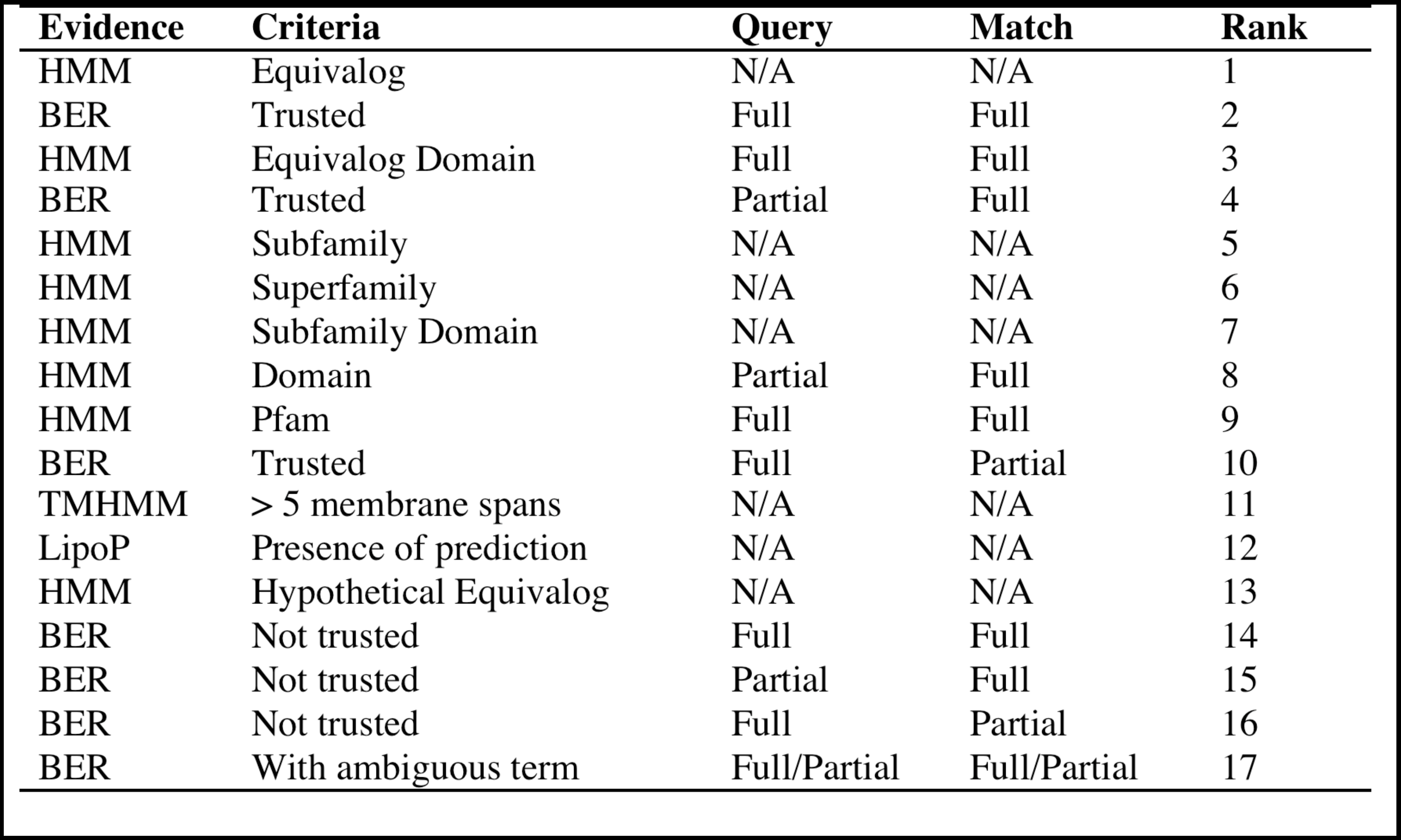
Protein A ~ Protein B ~ Protein C ~ Protein D

But, is Protein A similar to Protein D?

If not, a transitive annotation error has occurred.

To prevent, or at least minimize, such errors we require that a match protein be "trusted" if specific functional annotations are made from it.

prokaryotic protein functional prediction (pFunc)

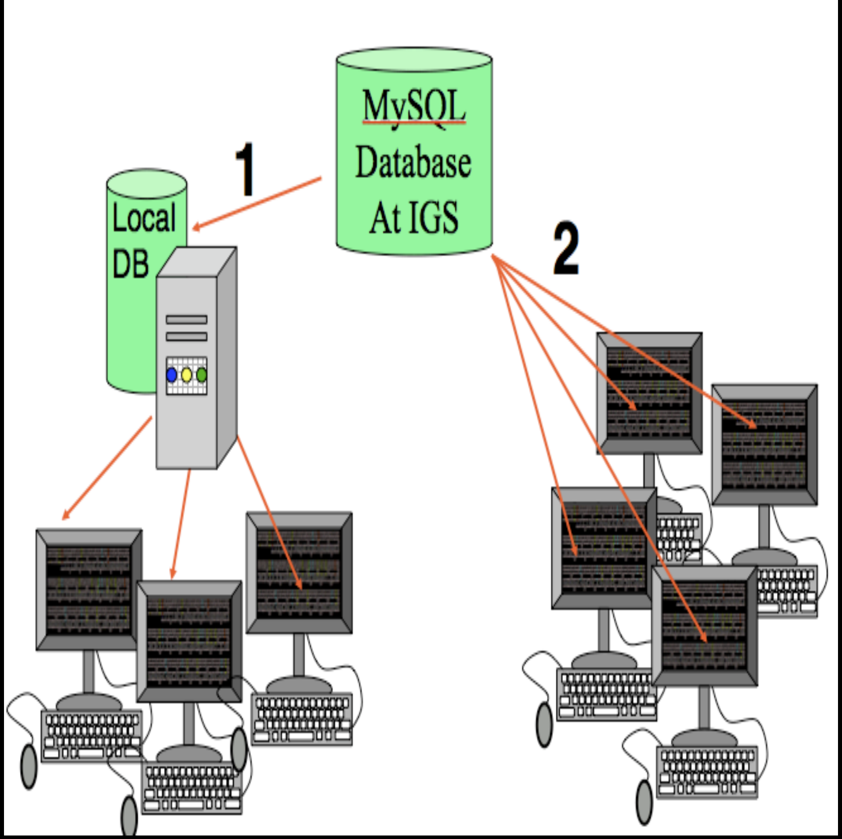


Evidence	Criteria	Query	Match	Rank
HMM	Equivalog	N/A	N/A	1
BER	Trusted	Full	Full	2
HMM	Equivalog Domain	Full	Full	3
BER	Trusted	Partial	Full	4
HMM	Subfamily	N/A	N/A	5
HMM	Superfamily	N/A	N/A	6
HMM	Subfamily Domain	N/A	N/A	7
HMM	Domain	Partial	Full	8
HMM	Pfam	Full	Full	9
BER	Trusted	Full	Partial	10
TMHMM	> 5 membrane spans	N/A	N/A	11
LipoP	Presence of prediction	N/A	N/A	12
HMM	Hypothetical Equivalog	N/A	N/A	13
BER	Not trusted	Full	Full	14
BER	Not trusted	Partial	Full	15
BER	Not trusted	Full	Partial	16
BER	With ambiguous term	Full/Partial	Full/Partial	17

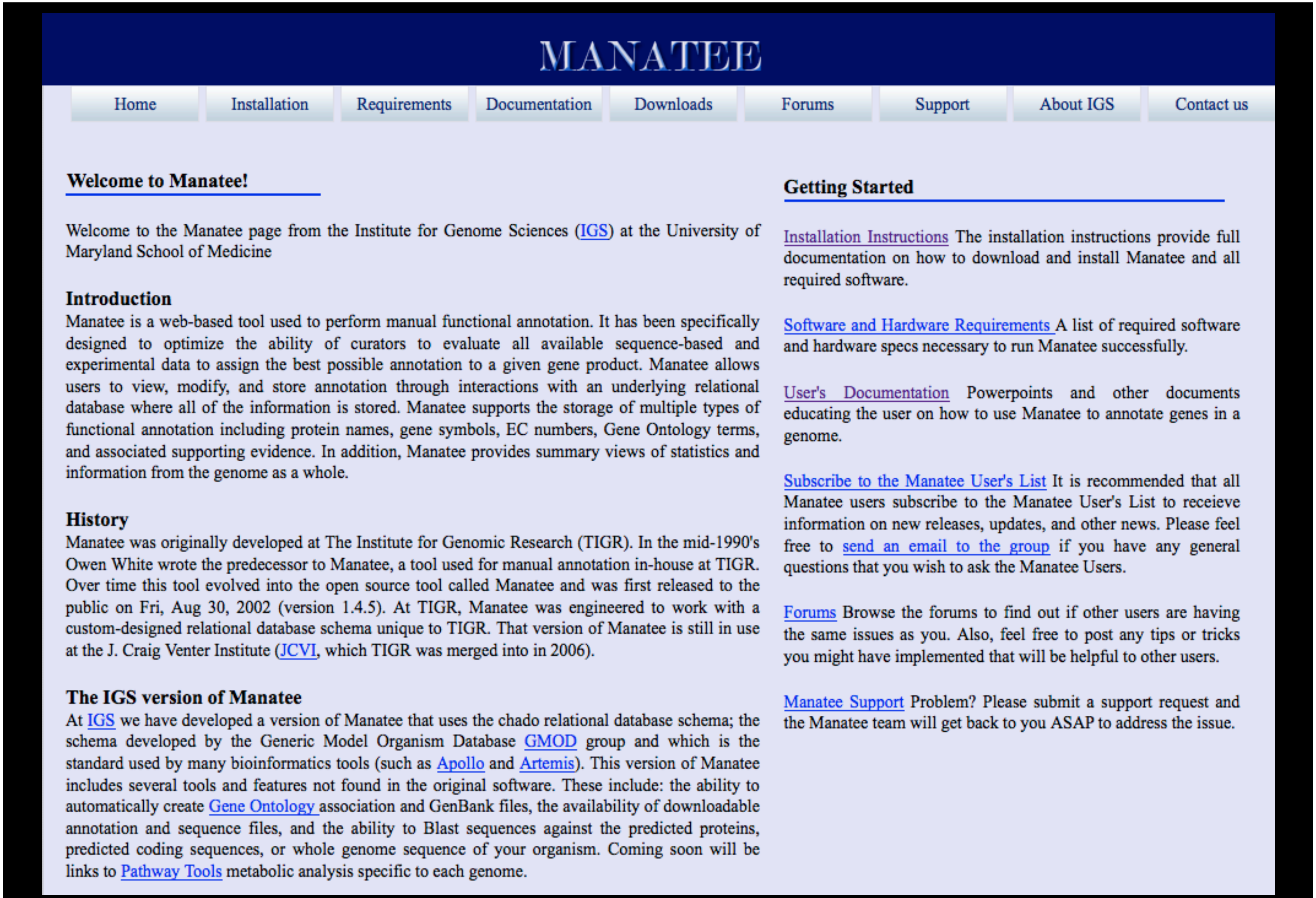
Protein names are adjusted to reflect functional confidence/specificity

- High confidence in specific function
  - “adenylosuccinate lyase” with EC/gene symbol
- General knowledge of function or subfamily
  - “carbohydrate kinase”, FGGY family
- Family/Domain membership
  - “cbbY family protein”
- Hypotheticals
  - “hypothetical protein
  - “conserved hypothetical protein”

Options for Data Access



- Option 1
  - We place a MySQl version of your database and files onto an ftp site. You download it and Manatee for local installation
- Option 2
  - Your database resides at IGS. We provide you a password-protected account to Manatee installed at IGS.
  - By far the most popular option.
- Option 3
  - File downloads
    - gff3
    - gbk
    - Simple tab-delimited with functional information
    - Multifasta protein/nucleotide



Welcome to Manatee!

Welcome to the Manatee page from the Institute for Genome Sciences (IGS) at the University of Maryland School of Medicine

Introduction

Manatee is a web-based tool used to perform manual functional annotation. It has been specifically designed to optimize the ability of curators to evaluate all available sequence-based and experimental data to assign the best possible annotation to a given gene product. Manatee allows users to view, modify, and store annotation through interactions with an underlying relational database where all of the information is stored. Manatee supports the storage of multiple types of functional annotation including protein names, gene symbols, EC numbers, Gene Ontology terms, and associated supporting evidence. In addition, Manatee provides summary views of statistics and information from the genome as a whole.

History

Manatee was originally developed at The Institute for Genomic Research (TIGR). In the mid-1990's Owen White wrote the predecessor to Manatee, a tool used for manual annotation in-house at TIGR. Over time this tool evolved into the open source tool called Manatee and was first released to the public on Fri, Aug 30, 2002 (version 1.4.5). At TIGR, Manatee was engineered to work with a custom-designed relational database schema unique to TIGR. That version of Manatee is still in use at the J. Craig Venter Institute (JCVI, which TIGR was merged into in 2006).

The IGS version of Manatee

At IGS we have developed a version of Manatee that uses the chado relational database schema; the schema developed by the Generic Model Organism Database GMOD group and which is the standard used by many bioinformatics tools (such as Apollo and Artemis). This version of Manatee includes several tools and features not found in the original software. These include: the ability to automatically create Gene Ontology association and GenBank files, the availability of downloadable annotation and sequence files, and the ability to Blast sequences against the predicted proteins, predicted coding sequences, or whole genome sequence of your organism. Coming soon will be links to Pathway Tools metabolic analysis specific to each genome.

Getting Started

Installation Instructions The installation instructions provide full documentation on how to download and install Manatee and all required software.

Software and Hardware Requirements A list of required software and hardware specs necessary to run Manatee successfully.

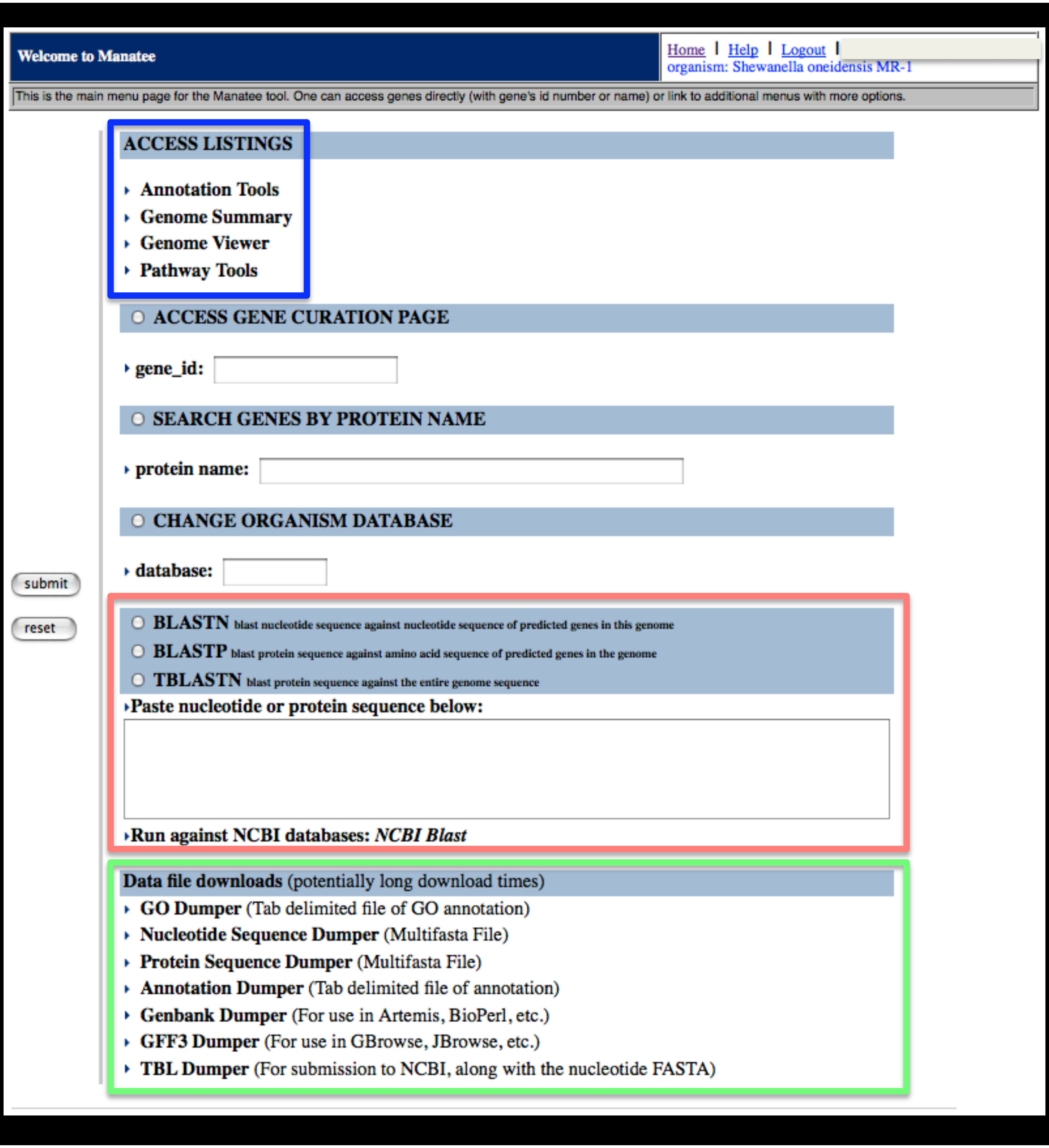
User's Documentation Powerpoints and other documents educating the user on how to use Manatee to annotate genes in a genome.

Subscribe to the Manatee User's List It is recommended that all Manatee users subscribe to the Manatee User's List to receieve information on new releases, updates, and other news. Please feel free to send an email to the group if you have any general questions that you wish to ask the Manatee Users.

Forums Browse the forums to find out if other users are having the same issues as you. Also, feel free to post any tips or tricks you might have implemented that will be helpful to other users.

Manatee Support Problem? Please submit a support request and the Manatee team will get back to you ASAP to address the issue.

manatee.sourceforge.net



Welcome to Manatee

Home | Help | Logout |
organism: Shewanella oneidensis MR-1

This is the main menu page for the Manatee tool. One can access genes directly (with gene's id number or name) or link to additional menus with more options.

ACCESS LISTINGS

- Annotation Tools
- Genome Summary
- Genome Viewer
- Pathway Tools

○ **ACCESS GENE CURATION PAGE**

- **gene_id:**

○ **SEARCH GENES BY PROTEIN NAME**

- **protein name:**

○ **CHANGE ORGANISM DATABASE**

- **database:**

submit

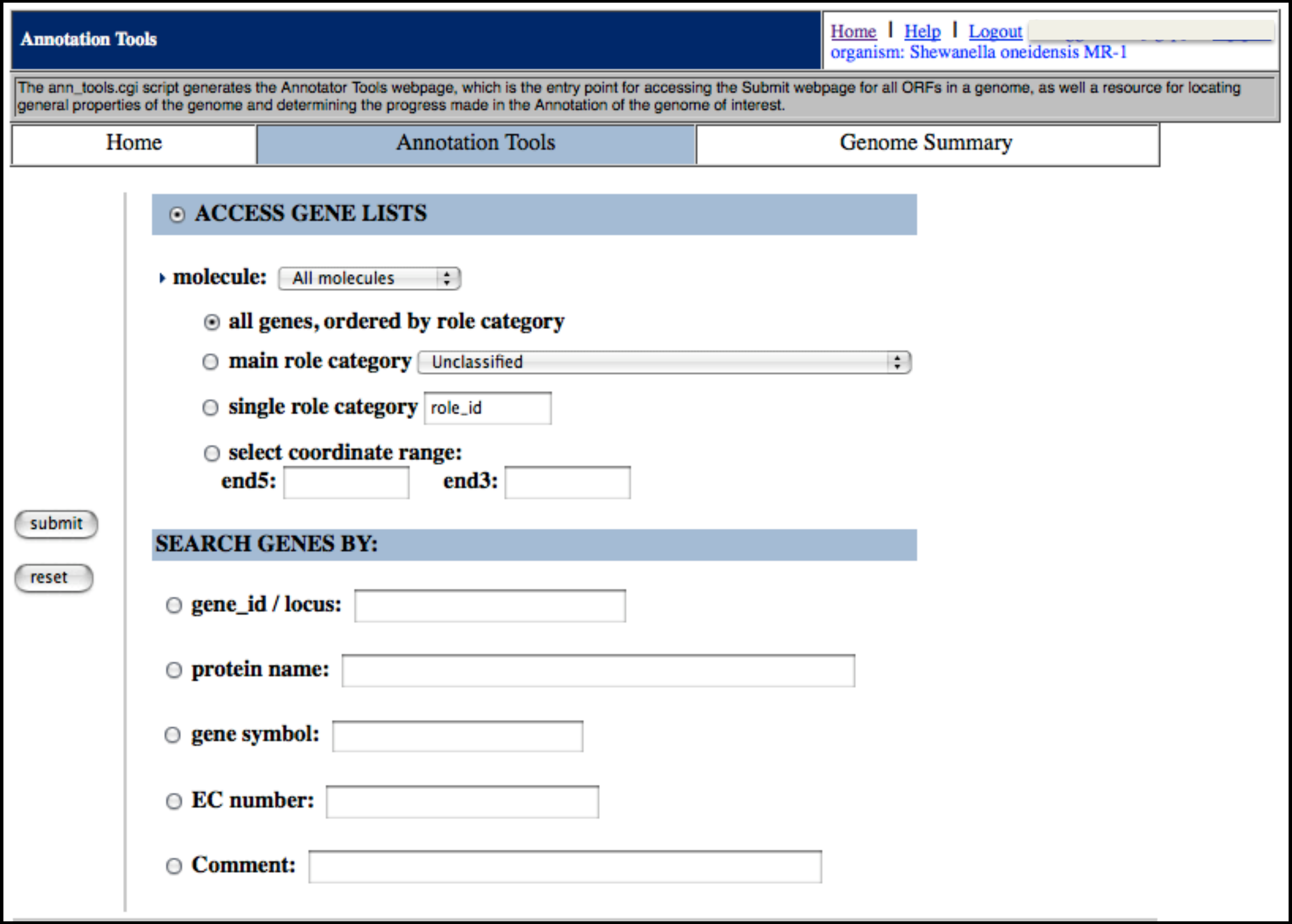
reset

○ **BLASTN** blast nucleotide sequence against nucleotide sequence of predicted genes in this genome
○ **BLASTP** blast protein sequence against amino acid sequence of predicted genes in the genome
○ **TBLASTN** blast protein sequence against the entire genome sequence

- **Paste nucleotide or protein sequence below:**
- **Run against NCBI databases:** ***NCBI Blast***

Data file downloads (potentially long download times)

- **GO Dumper** (Tab delimited file of GO annotation)
- **Nucleotide Sequence Dumper** (Multifasta File)
- **Protein Sequence Dumper** (Multifasta File)
- **Annotation Dumper** (Tab delimited file of annotation)
- **Genbank Dumper** (For use in Artemis, BioPerl, etc.)
- **GFF3 Dumper** (For use in GBrowse, JBrowse, etc.)
- **TBL Dumper** (For submission to NCBI, along with the nucleotide FASTA)



Annotation Tools

Home | Help | Logout [redacted]
organism: Shewanella oneidensis MR-1

The ann_tools.cgi script generates the Annotator Tools webpage, which is the entry point for accessing the Submit webpage for all ORFs in a genome, as well a resource for locating general properties of the genome and determining the progress made in the Annotation of the genome of interest.

● ACCESS GENE LISTS

- molecule: All molecules

● all genes, ordered by role category

○ main role category: Unclassified

○ single role category: role_id

○ select coordinate range:
end5: end3:

SEARCH GENES BY:

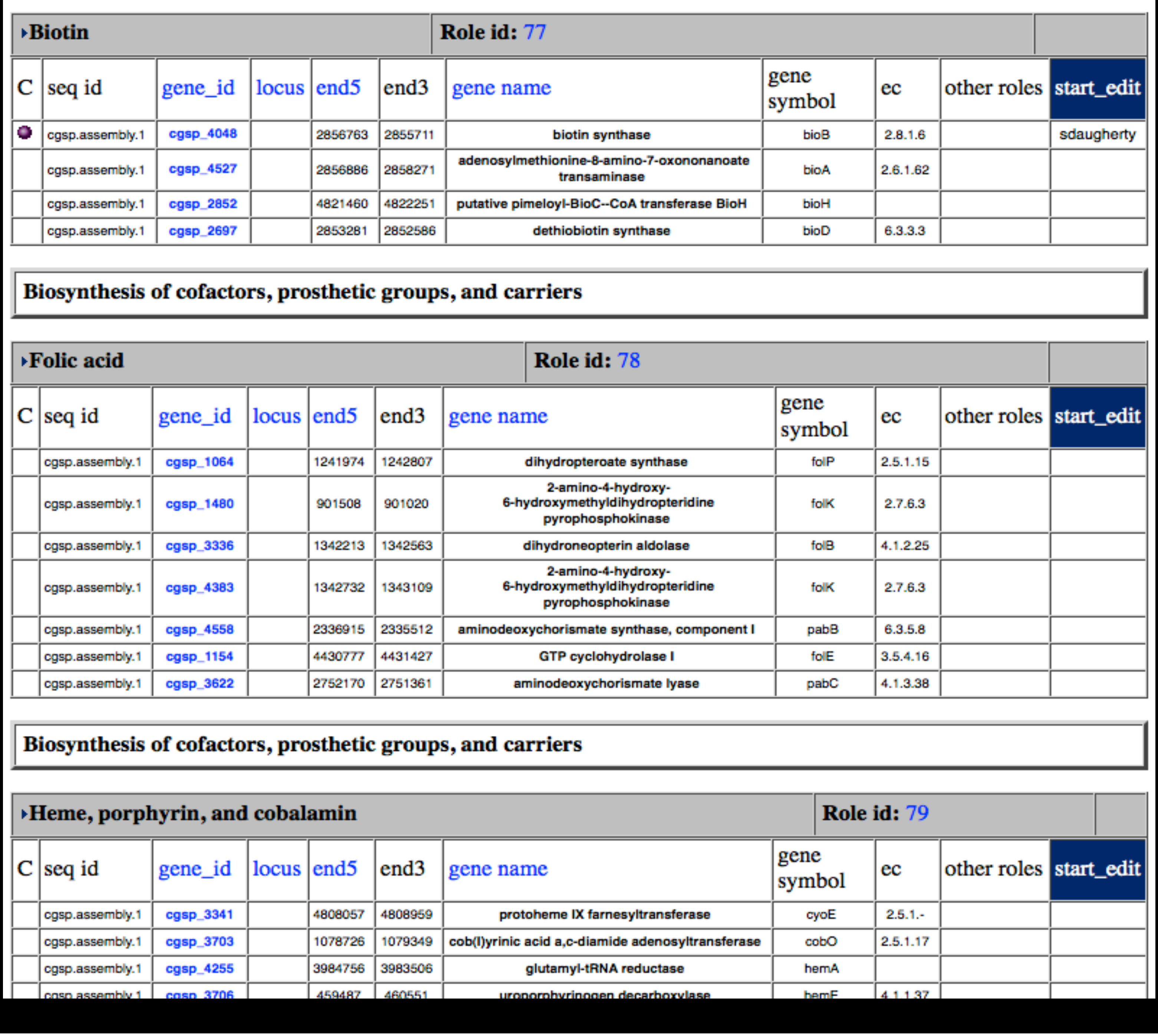
○ gene_id / locus:

○ protein name:

○ gene symbol:

○ EC number:

○ Comment:



‣Biotin | Role id: 77

C	seq id	gene_id	locus	end5	end3	gene name	gene symbol	ec	other roles	start_edit
●	cgsp.assembly.1	cgsp_4048		2856763	2855711	biotin synthase	bioB	2.8.1.6		sdaugherty
	cgsp.assembly.1	cgsp_4527		2856886	2858271	adenosylmethionine-8-amino-7-oxononanoate transaminase	bioA	2.6.1.62		
	cgsp.assembly.1	cgsp_2852		4821460	4822251	putative pimeloyl-BioC--CoA transferase BioH	bioH			
	cgsp.assembly.1	cgsp_2697		2853281	2852586	dethiobiotin synthase	bioD	6.3.3.3		

Biosynthesis of cofactors, prosthetic groups, and carriers

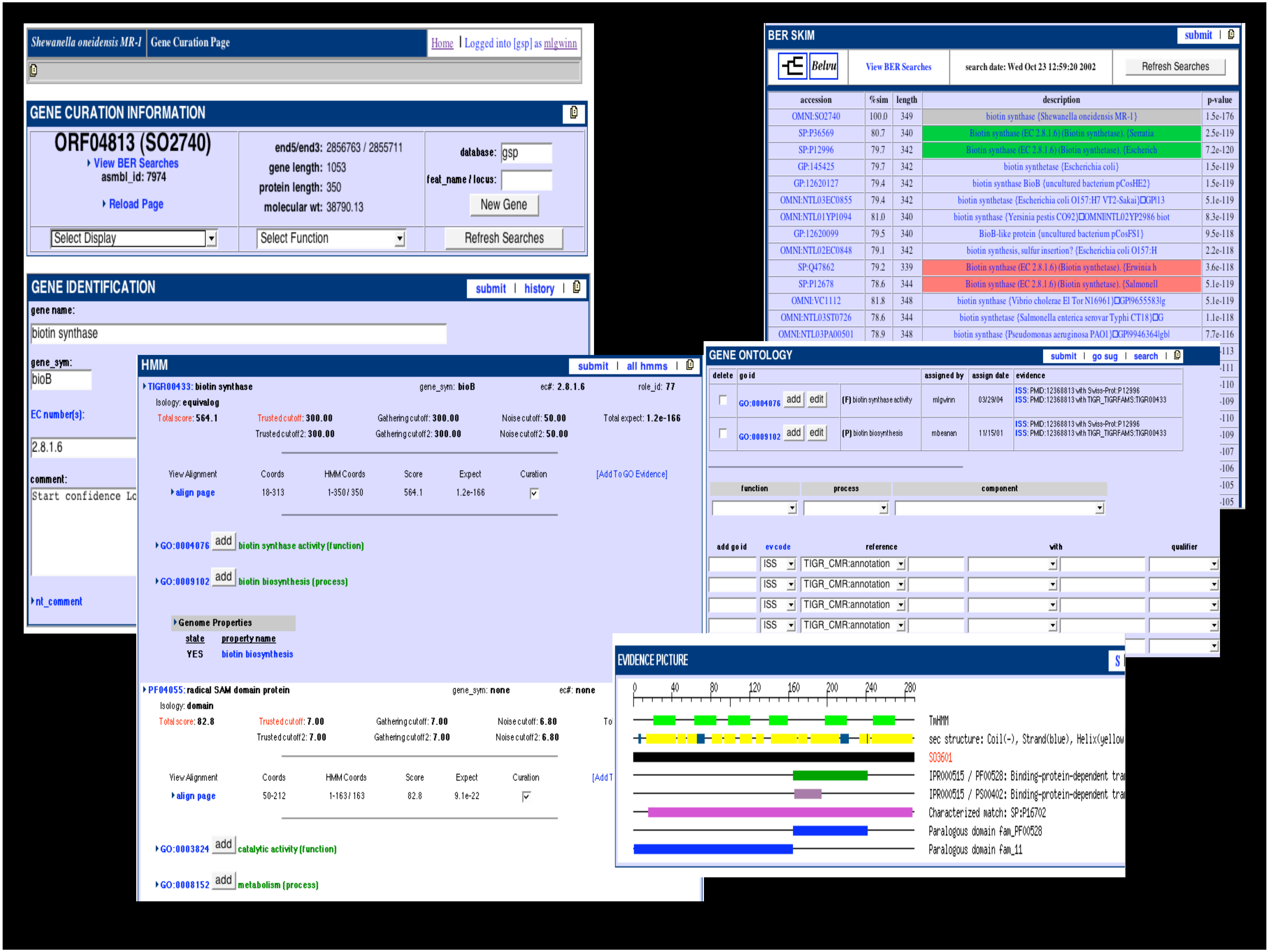
‣Folic acid | Role id: 78

C	seq id	gene_id	locus	end5	end3	gene name	gene symbol	ec	other roles	start_edit
	cgsp.assembly.1	cgsp_1064		1241974	1242807	dihydropteroate synthase	folP	2.5.1.15		
	cgsp.assembly.1	cgsp_1480		901508	901020	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	folK	2.7.6.3		
	cgsp.assembly.1	cgsp_3336		1342213	1342563	dihydroneopterin aldolase	folB	4.1.2.25		
	cgsp.assembly.1	cgsp_4383		1342732	1343109	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	folK	2.7.6.3		
	cgsp.assembly.1	cgsp_4558		2336915	2335512	aminodeoxychorismate synthase, component I	pabB	6.3.5.8		
	cgsp.assembly.1	cgsp_1154		4430777	4431427	GTP cyclohydrolase I	folE	3.5.4.16		
	cgsp.assembly.1	cgsp_3622		2752170	2751361	aminodeoxychorismate lyase	pabC	4.1.3.38		

Biosynthesis of cofactors, prosthetic groups, and carriers

‣Heme, porphyrin, and cobalamin | Role id: 79

C	seq id	gene_id	locus	end5	end3	gene name	gene symbol	ec	other roles	start_edit
	cgsp.assembly.1	cgsp_3341		4808057	4808959	protoheme IX farnesyltransferase	cyoE	2.5.1.-		
	cgsp.assembly.1	cgsp_3703		1078726	1079349	cob(I)yrinic acid a,c-diamide adenosyltransferase	cobO	2.5.1.17		
	cgsp.assembly.1	cgsp_4255		3984756	3983506	glutamyl-tRNA reductase	hemA			
	cgsp.assembly.1	cgsp_3706		459487	460551	uroporphyrinogen decarboxylase	hemE	4.1.1.37		



Shewanella oneidensis MR-1 | Gene Curation Page | Home | Logged into [gsp] as mlgwinn

GENE CURATION INFORMATION

ORF04813 (SO2740)
View BER Searches
asmbl_id: 7974
Reload Page

end5/end3: 2856763 / 2855711
gene length: 1053
protein length: 350
molecular wt: 38790.13

database: gsp
feat_name / locus:
New Gene

Select Display | Select Function | Refresh Searches

GENE IDENTIFICATION

submit | history

gene name: biotin synthase

gene_sym: bioB

EC number(s): 2.8.1.6

comment: Start confidence Lo

nt_comment

HMM

submit | all hmms

TIGR00433: biotin synthase gene_sym: **bioB** ec#: **2.8.1.6** role_id: **77**

Isology: **equivalog**

Total score: **564.1** Trusted cutoff: **300.00** Gathering cutoff: **300.00** Noise cutoff: **50.00** Total expect: **1.2e-166**
Trusted cutoff2: **300.00** Gathering cutoff2: **300.00** Noise cutoff2: **50.00**

View Alignment	Coords	HMM Coords	Score	Expect	Curation	[Add To GO Evidence]
align page	18-313	1-350 / 350	564.1	1.2e-166	☑	

GO:0004076 add biotin synthase activity (function)

GO:0009102 add biotin biosynthesis (process)

Genome Properties

state	property name
YES	biotin biosynthesis

PF04055: radical SAM domain protein gene_sym: **none** ec#: **none**

Isology: **domain**

Total score: **82.8** Trusted cutoff: **7.00** Gathering cutoff: **7.00** Noise cutoff: **6.80** To
Trusted cutoff2: **7.00** Gathering cutoff2: **7.00** Noise cutoff2: **6.80**

View Alignment	Coords	HMM Coords	Score	Expect	Curation	[Add T
align page	50-212	1-163 / 163	82.8	9.1e-22	☑	

GO:0003824 add catalytic activity (function)

GO:0008152 add metabolism (process)

BER SKIM

submit

Belvu | View BER Searches | search date: Wed Oct 23 12:59:20 2002 | Refresh Searches

accession	%sim	length	description	p-value
OMNI:SO2740	100.0	349	biotin synthase {Shewanella oneidensis MR-1}	1.5e-176
SP:P36569	80.7	340	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Serratia	2.5e-119
SP:P12996	79.7	342	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Escherich	7.2e-120
GP:145425	79.7	342	biotin synthetase {Escherichia coli}	1.5e-119
GP:12620127	79.4	342	biotin synthase BioB {uncultured bacterium pCosHE2}	1.5e-119
OMNI:NTL03EC0855	79.4	342	biotin synthetase {Escherichia coli O157:H7 VT2-Sakai}□GP|13	5.1e-119
OMNI:NTL01YP1094	81.0	340	biotin synthase {Yersinia pestis CO92}□OMNI|NTL02YP2986 biot	8.3e-119
GP:12620099	79.5	340	BioB-like protein {uncultured bacterium pCosFS1}	9.5e-118
OMNI:NTL02EC0848	79.1	342	biotin synthesis, sulfur insertion? {Escherichia coli O157:H	2.2e-118
SP:Q47862	79.2	339	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Erwinia h	3.6e-118
SP:P12678	78.6	344	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Salmonell	5.1e-119
OMNI:VC1112	81.8	348	biotin synthase {Vibrio cholerae El Tor N16961}□GP|9655583|g	5.1e-119
OMNI:NTL03ST0726	78.6	344	biotin synthetase {Salmonella enterica serovar Typhi CT18}□G	1.1e-118
OMNI:NTL03PA00501	78.9	348	biotin synthase {Pseudomonas aeruginosa PAO1}□GP|9946364|gb|	7.7e-116
				-113
				-111
				-110
				-109
				-110
				-109
				-107
				-106
				-105
				-105

GENE ONTOLOGY

submit | go sug | search

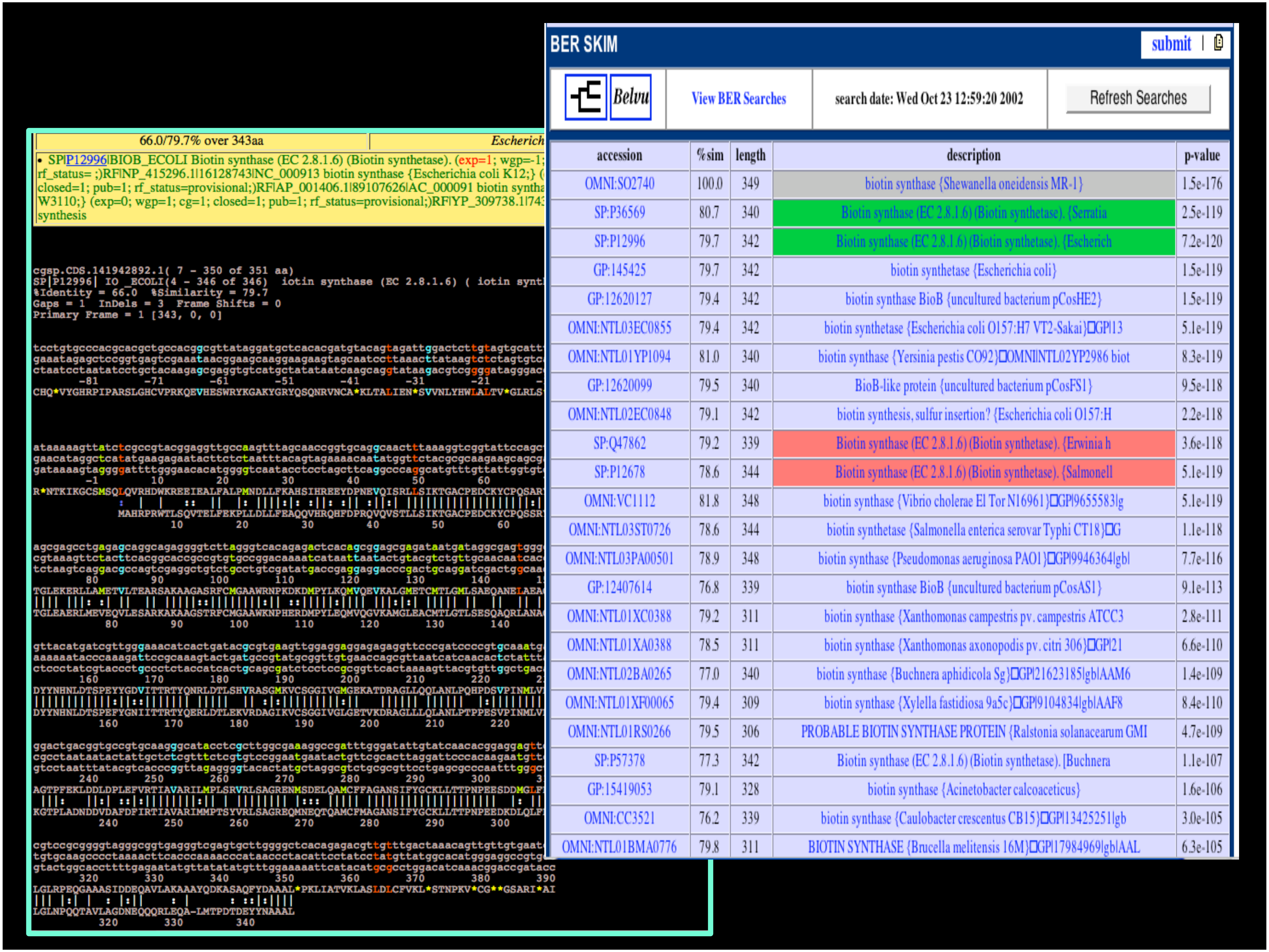
delete	go id			assigned by	assign date	evidence
□	GO:0004076	add edit	(F) biotin synthase activity	mlgwinn	03/29/04	ISS: PMID:12368813 with Swiss-Prot:P12996 ISS: PMID:12368813 with TIGR_TIGRFAMS:TIGR00433
□	GO:0009102	add edit	(P) biotin biosynthesis	mbeanan	11/15/01	ISS: PMID:12368813 with Swiss-Prot:P12996 ISS: PMID:12368813 with TIGR_TIGRFAMS:TIGR00433

function	process	component

add go id	ev code	reference		with		qualifier
	ISS	TIGR_CMR:annotation				
	ISS	TIGR_CMR:annotation				
	ISS	TIGR_CMR:annotation				
	ISS	TIGR_CMR:annotation				

EVIDENCE PICTURE

TmHMM
sec structure: Coil(-), Strand(blue), Helix(yellow
SO3601
IPR000515 / PF00528: Binding-protein-dependent tra
IPR000515 / PS00402: Binding-protein-dependent tra
Characterized match: SP:P16702
Paralogous domain fam_PF00528
Paralogous domain fam_11



66.0/79.7% over 343aa | *Escherich*

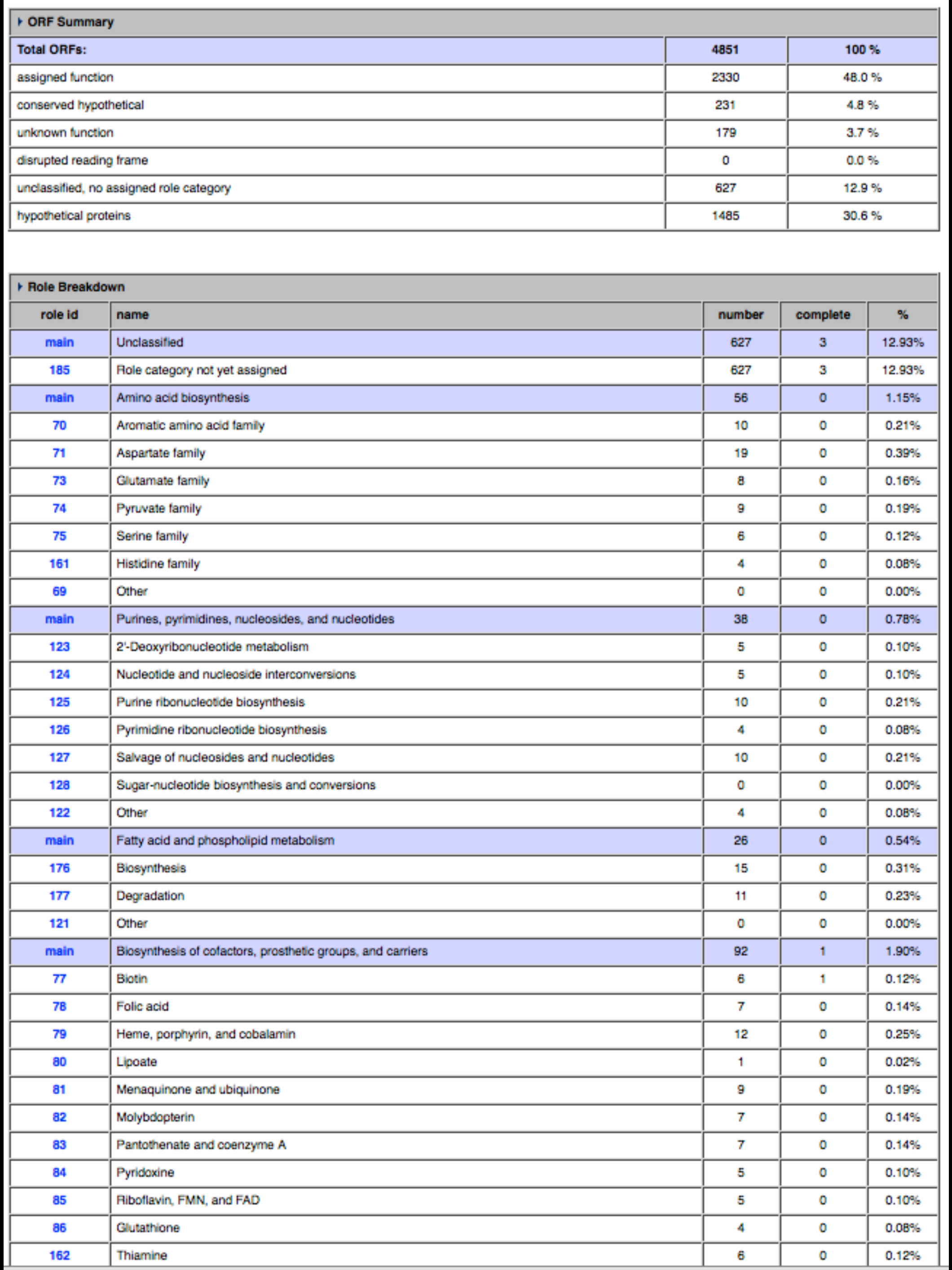
• SP|P12996|BIOB_ECOLI Biotin synthase (EC 2.8.1.6) (Biotin synthetase). (exp=1; wgp=-1;
rf_status= ;)RF|NP_415296.1|16128743|NC_000913 biotin synthase {Escherichia coli K12;} (
closed=1; pub=1; rf_status=provisional;)RF|AP_001406.1|89107626|AC_000091 biotin syntha
W3110;} (exp=0; wgp=1; cg=1; closed=1; pub=1; rf_status=provisional;)RF|YP_309738.1|743
synthesis

```
cgsp.CDS.141942892.1( 7 - 350 of 351 aa)
SP|P12996| IO _ECOLI(4 - 346 of 346)  iotin synthase (EC 2.8.1.6) ( iotin synt
%Identity = 66.0  %Similarity = 79.7
Gaps = 1  InDels = 3  Frame Shifts = 0
Primary Frame = 1 [343, 0, 0]


tcctgtgcccacgcacgctgccacggcgttataggatgctcacacgatgtacagtagattggactcttgtagtgcatt
gaaatagagctccggtgagtcgaaataacggaagcaaggaagaagtagcaatccttaaacttataagtctctagtgtc
ctaatcctaatatcctgctacaagagcgaggtgtcatgctatataatcaagcaggtataagacgtcgggatagggac
       -81       -71       -61       -51       -41       -31       -21       -
CHQ*VYGHRPIPARSLGHCVPRKQEVHESWRYKGAKYGRYQSQNRVNCA*KLTALIEN*SVVNLYHWLALTV*GLRLS



ataaaaagttatctcgccgtacggaggttgccaagtttagcaaccggtgcaggcaactttaaaggtcggtattccagc
gaacataggctcatatgaagagaatacttctctaatttacagtagaaaacaatatggttctacgcgcaagaagcagcg
gataaaagtaggggatttgggaacacatggggtcaatacctcctagcttcaggcccaggcatgtttgttattggtgt
          -1        10        20        30        40        50        60
R*NTKIKGCSMSQLQVRHDWKREEIEALFALPMNDLLFKAHSIHREEYDPNEVQISRLLSIKTGACPEDCKYCPQSAR
             :  |  |   ::  ||  |: ||||:|: :||: :|| :||:| ||||||||||||||||||||:|
             MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVSTLLSIKTGACPEDCKYCPQSSR
                      10        20        30        40        50        60

agcgagcctgagagcaggcagaggggtcttagggtcacagagactcacagcggagcgagataatgataggcgagtggg
cgtaaagttctacttcacggcaccgccgtgtgccggacaaaatcataattaatactgtacgtctgttgcaacaatcac
tctaagtcaggacgccagtcgaggctgtctgcctgtcgatatgaccgaggaggacccgactgcaggatcgactggcaa
        80        90       100       110       120       130       140       15
TGLEKERLLAMETVLTEARSAKAAGASRFCMGAAWRNPKDKDMPYLKQMVQEVKALGMETCMTLGMLSAEQANELAEA
|||| |||: :| ||  || |||||::||||||||:|| ::||||:|||| |||:|:| ||||| ||   ||   || |
TGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAWKNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANA
           80        90       100       110       120       130       140

gttacatgatcgttgggaaacatcactgatacgcgtgaagttggaggaggagagaggttcccgatccccgtgcaaatg
aaaaaatacccaaagattccgcaaagtactgatgccgtatgcggttgtgaaccagcgttaatcatcaacactctattt
ctccctatcgtaccctgccctctaccatcactgcagcgatctcctccgcggttcactaaaagttacgtgttggctgac
       160       170       180       190       200       210       220       2
DYYNHNLDTSPEYYGDVITTRTYQNRLDTLSHVRASGMKVCSGGIVGMGEKATDRAGLLQQLANLPQHPDSVPINMLV
|||||||||||||:||::||||||| |||||  || :|:||||||||||:||    ||||||  ||||||   |:|||||
DYYNHNLDTSPEFYGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPESVPINMLV
          160       170       180       190       200       210       220

ggactgacggtgccgtgcaaggcatacctcgcttggcgaaaggccgatttgggatattgtatcaacacggaggagtt
cgcctaataatactattgctctcgtttctcgtgtccggaatgaatactgttcgcacttaggattcccacaagaatgtt
gtcctaatttatacgtcacccggttagaggggtacactatgctaggcgtctgcgcgttcctgagcgcccaatttgggc
       240       250       260       270       280       290       300       3
AGTPFEKLDDLDPLEFVRTIAVARILMPLSRVRLSAGRENMSDELQAMCFFAGANSIFYGCKLLTTPNPEESDDMGLF
 |||:    ||:| ::|:||||||||:|| | |||||||| |::: ||||| ||||||||||||||||||||   |: ||
KGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMCFMAGANSIFYGCKLLTTPNPEEDKDLQLF
          240       250       260       270       280       290       300

cgtccgcggggtagggcggtgagggtcgagtgcttggggctcacagagacgttgtttgactaaacagttgttgtgaat
tgtgcaagcccctaaaacttcacccaaaacccataaccctacattcctatcctatgttatggcacatgggaggccgtg
gtactggcacctttttgagaatatgttatatatgtttggaaaaattcatacatgcgcctggacatcaaacggaccgatacc
       320       330       340       350       360       370       380       390
LGLRPEQGAAASIDDEQAVLAKAAAYQDKASAQFYDAAAL*PKLIATVKLASLDLCFVKL*STNPKV*CG**GSARI*AI
||| |:| |   : |:||    : |       : ::|:||||
LGLNPQQTAVLAGDNEQQQRLEQA-LMTPDTDEYYNAAAL
          320       330       340
```

BER SKIM

search date: Wed Oct 23 12:59:20 2002

accession	%sim	length	description	p-value
OMNI:SO2740	100.0	349	biotin synthase {Shewanella oneidensis MR-1}	1.5e-176
SP:P36569	80.7	340	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Serratia	2.5e-119
SP:P12996	79.7	342	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Escherich	7.2e-120
GP:145425	79.7	342	biotin synthetase {Escherichia coli}	1.5e-119
GP:12620127	79.4	342	biotin synthase BioB {uncultured bacterium pCosHE2}	1.5e-119
OMNI:NTL03EC0855	79.4	342	biotin synthetase {Escherichia coli O157:H7 VT2-Sakai}□GP|13	5.1e-119
OMNI:NTL01YP1094	81.0	340	biotin synthase {Yersinia pestis CO92}□OMNI|NTL02YP2986 biot	8.3e-119
GP:12620099	79.5	340	BioB-like protein {uncultured bacterium pCosFS1}	9.5e-118
OMNI:NTL02EC0848	79.1	342	biotin synthesis, sulfur insertion? {Escherichia coli O157:H	2.2e-118
SP:Q47862	79.2	339	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Erwinia h	3.6e-118
SP:P12678	78.6	344	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Salmonell	5.1e-119
OMNI:VC1112	81.8	348	biotin synthase {Vibrio cholerae El Tor N16961}□GP|9655583|g	5.1e-119
OMNI:NTL03ST0726	78.6	344	biotin synthetase {Salmonella enterica serovar Typhi CT18}□G	1.1e-118
OMNI:NTL03PA00501	78.9	348	biotin synthase {Pseudomonas aeruginosa PAO1}□GP|9946364|gb|	7.7e-116
GP:12407614	76.8	339	biotin synthase BioB {uncultured bacterium pCosAS1}	9.1e-113
OMNI:NTL01XC0388	79.2	311	biotin synthase {Xanthomonas campestris pv. campestris ATCC3	2.8e-111
OMNI:NTL01XA0388	78.5	311	biotin synthase {Xanthomonas axonopodis pv. citri 306}□GP|21	6.6e-110
OMNI:NTL02BA0265	77.0	340	biotin synthase {Buchnera aphidicola Sg}□GP|21623185|gb|AAM6	1.4e-109
OMNI:NTL01XF00065	79.4	309	biotin synthase {Xylella fastidiosa 9a5c}□GP|9104834|gb|AAF8	8.4e-110
OMNI:NTL01RS0266	79.5	306	PROBABLE BIOTIN SYNTHASE PROTEIN {Ralstonia solanacearum GMI	4.7e-109
SP:P57378	77.3	342	Biotin synthase (EC 2.8.1.6) (Biotin synthetase). [Buchnera	1.1e-107
GP:15419053	79.1	328	biotin synthase {Acinetobacter calcoaceticus}	1.6e-106
OMNI:CC3521	76.2	339	biotin synthase {Caulobacter crescentus CB15}□GP|13425251|gb	3.0e-105
OMNI:NTL01BMA0776	79.8	311	BIOTIN SYNTHASE {Brucella melitensis 16M}□GP|17984969|gb|AAL	6.3e-105

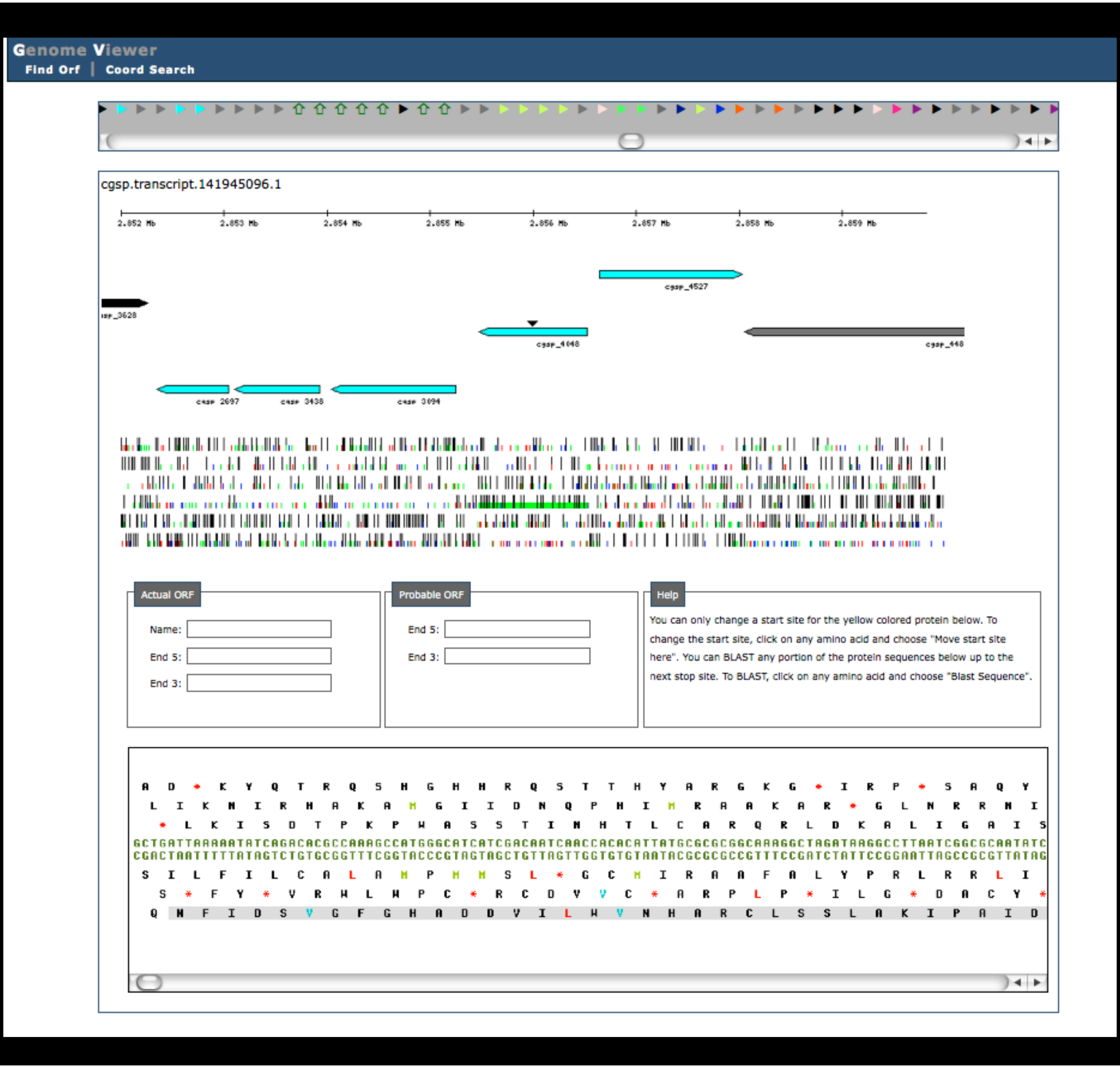


▸ ORF Summary

Total ORFs:	4851	100 %
assigned function	2330	48.0 %
conserved hypothetical	231	4.8 %
unknown function	179	3.7 %
disrupted reading frame	0	0.0 %
unclassified, no assigned role category	627	12.9 %
hypothetical proteins	1485	30.6 %

▸ Role Breakdown

role id	name	number	complete	%
main	Unclassified	627	3	12.93%
185	Role category not yet assigned	627	3	12.93%
main	Amino acid biosynthesis	56	0	1.15%
70	Aromatic amino acid family	10	0	0.21%
71	Aspartate family	19	0	0.39%
73	Glutamate family	8	0	0.16%
74	Pyruvate family	9	0	0.19%
75	Serine family	6	0	0.12%
161	Histidine family	4	0	0.08%
69	Other	0	0	0.00%
main	Purines, pyrimidines, nucleosides, and nucleotides	38	0	0.78%
123	2'-Deoxyribonucleotide metabolism	5	0	0.10%
124	Nucleotide and nucleoside interconversions	5	0	0.10%
125	Purine ribonucleotide biosynthesis	10	0	0.21%
126	Pyrimidine ribonucleotide biosynthesis	4	0	0.08%
127	Salvage of nucleosides and nucleotides	10	0	0.21%
128	Sugar-nucleotide biosynthesis and conversions	0	0	0.00%
122	Other	4	0	0.08%
main	Fatty acid and phospholipid metabolism	26	0	0.54%
176	Biosynthesis	15	0	0.31%
177	Degradation	11	0	0.23%
121	Other	0	0	0.00%
main	Biosynthesis of cofactors, prosthetic groups, and carriers	92	1	1.90%
77	Biotin	6	1	0.12%
78	Folic acid	7	0	0.14%
79	Heme, porphyrin, and cobalamin	12	0	0.25%
80	Lipoate	1	0	0.02%
81	Menaquinone and ubiquinone	9	0	0.19%
82	Molybdopterin	7	0	0.14%
83	Pantothenate and coenzyme A	7	0	0.14%
84	Pyridoxine	5	0	0.10%
85	Riboflavin, FMN, and FAD	5	0	0.10%
86	Glutathione	4	0	0.08%
162	Thiamine	6	0	0.12%



cgsp.transcript.141945096.1

Name:
End 5:
End 3:

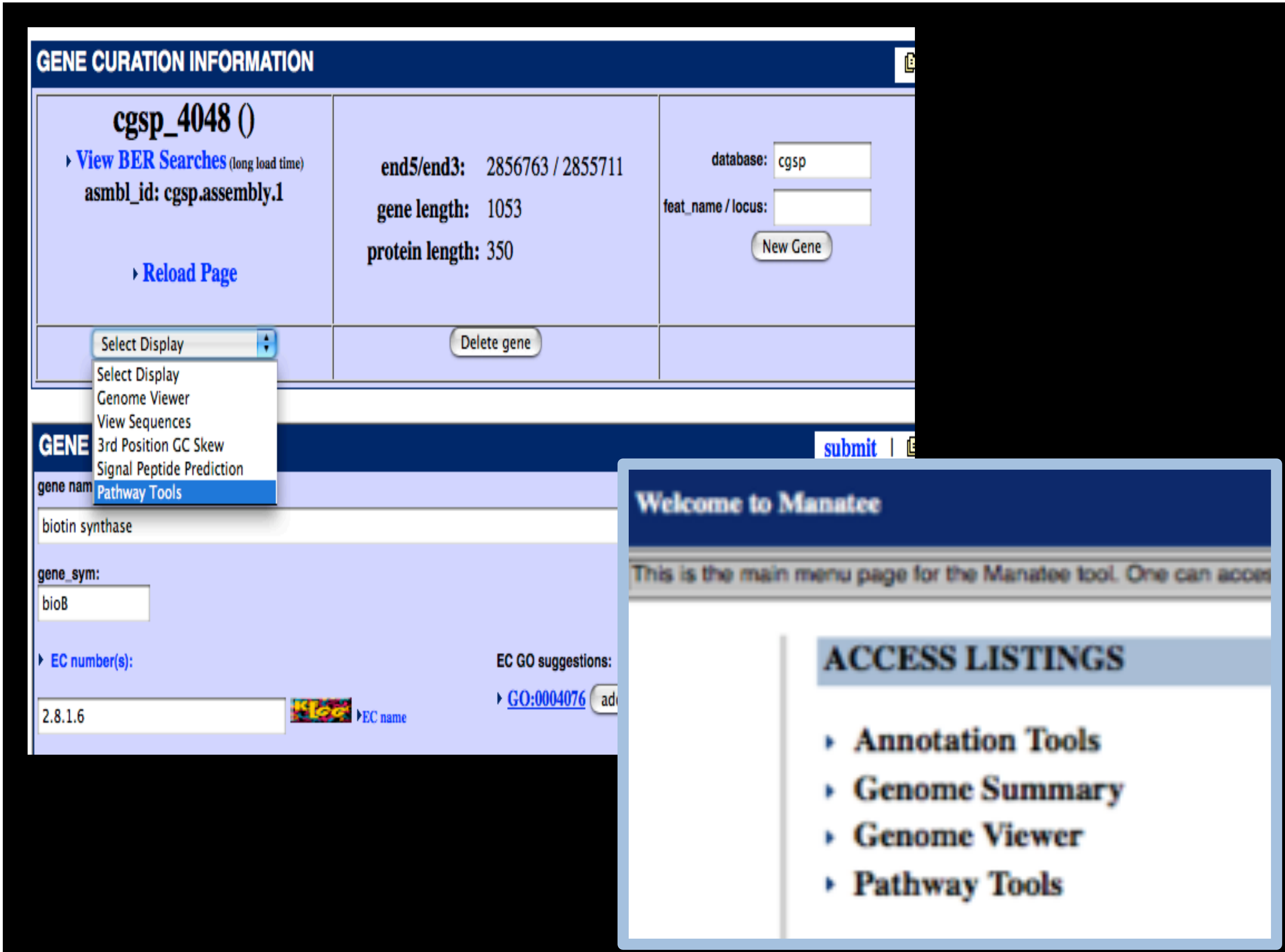
End 5:
End 3:

You can only change a start site for the yellow colored protein below. To change the start site, click on any amino acid and choose "Move start site here". You can BLAST any portion of the protein sequences below up to the next stop site. To BLAST, click on any amino acid and choose "Blast Sequence".

```
A  D  *  K  Y  Q  T  R  Q  S  H  G  H  H  R  Q  S  T  T  H  Y  A  R  G  K  G  *  I  R  P  *  S  A  Q  Y
 L  I  K  N  I  R  H  A  K  A  M  G  I  I  D  N  Q  P  H  I  M  R  A  A  K  A  R  *  G  L  N  R  R  N  I
  *  L  K  I  S  D  T  P  K  P  W  A  S  S  T  I  N  H  T  L  C  A  R  Q  R  L  D  K  A  L  I  G  A  I  S
GCTGATTAAAAATATCAGACACGCCAAAGCCATGGGCATCATCGACAATCAACCACACATTATGCGCGCGGCAAAGGCTAGATAAGGCCTTAATCGGCGCAATATC
CGACTAATTTTTATAGTCTGTGCGGTTTCGGTACCCGTAGTAGCTGTTAGTTGGTGTGTAATACGCGCGCCGTTTCCGATCTATTCCGGAATTAGCCGCGTTATAG
S  I  L  F  I  L  C  A  L  A  M  P  M  M  S  L  *  G  C  M  I  R  A  A  F  A  L  Y  P  R  L  R  R  L  I
  S  *  F  Y  *  V  R  W  L  W  P  C  *  R  C  D  V  V  C  *  A  R  P  L  P  *  I  L  G  *  D  A  C  Y  *
 Q  N  F  I  D  S  V  G  F  G  H  A  D  D  V  I  L  W  V  N  H  A  R  C  L  S  S  L  A  K  I  P  A  I  D
```

Pathway Tools

- All AE genomes now get Pathway Tools analysis
- A PGDB is created for each genome
- The PGDB is Available to the users via protected web site
- We are just beginning to form links between Manatee and the PGDBs



GENE CURATION INFORMATION

cgsp_4048 ()

View BER Searches (long load time)

asmbl_id: cgsp.assembly.1

Reload Page

end5/end3: 2856763 / 2855711

gene length: 1053

protein length: 350

database: cgsp

feat_name / locus:

gene nam

biotin synthase

gene_sym:

bioB

EC number(s):

2.8.1.6

EC GO suggestions:

GO:0004076

Welcome to Manatee

This is the main menu page for the Manatee tool. One can acces

ACCESS LISTINGS

- Annotation Tools
- Genome Summary
- Genome Viewer
- Pathway Tools

Future directions

- We are working on grant renewal now
  - Just entered our 4th and last year of the current grant

- We plan several more enhancements
  - more search options in Manatee
  - More customizable download/viewing options
  - Incorporation of new datatypes such as RNAseq

- Integration with other tools
  - Artemis
  - Apollo
  - IGS resources
    - Sybil
    - Mummer-remap

Future directions of Annotation Engine and Pathway Tools

- Communication between Manatee/PGDBs
  - – Lists of/links to pathways on Manatee GCPs
  - – Links to pathways from Manatee GCPs
- Use PT analysis to inform automatic annotation process in an iterative fashion
- Changes in Manatee propagate to PGDB and back again, automatic refresh of pathway predictions.

http://gscid.igs.umaryland.edu

University of Maryland School of Medicine
INSTITUTE FOR GENOME SCIENCES

GSCID

Genomic Sequencing Center for Infectious Diseases

The Institute for Genome Sciences (IGS) at the University of Maryland School of Medicine (UMSOM) has been awarded a five-year contract by the National Institute of Allergy and Infectious Diseases (NIAID) at NIH to establish a new Genomic Sequencing Center for Infectious Disease (GSCID).

The Genomic Sequencing Center for Infectious Disease will provide researchers with rapid and cost-efficient production of high-quality genome sequences of NIAID Category A-C priority pathogens, related organisms, clinical isolates, c ely related species, and inve ebrate vectors of infectious di eases and microorganisms responsible for ing and re-emerging inf s diseases. The GSCID dresses the need for sequencing microorgani invertebrate vector ase that are co gents of bioterrorism and/or high priority pa t could be p concerns and new analytical tools generated by the GS ared pub re icity, drug resistance, transmission and ment

IGS Genomics Workshop - 4 times per year

http://ae/cgi/workshop_info.cgi

Topics

-sequencing

-gene finding (prok and euk)

-functional annotation

-Gene Ontology

-Manatee demo and hands-on

-comparative genomics, Sybil demo

-Artemis demo

-expression analysis

-metagenomics

-Human Microbiome Project

-databases

-pipeline management

Please check out
the IGS careers page at:
http://www.igs.umaryland.edu

Acknowledgements

- Kevin Galens, Joshua Orvis
- Todd Creasy
- Sean Daugherty, Heather Creasy
- Jennifer Wortman, Anup Mahurkar
- Tanja Davidsen, Owen White
- Especially:

National Institute of
General Medical Sciences
One of the National Institutes of Health

for funding this project