

# A Guide to



Michelle Gwinn  
August, 2005

# Table of Contents (for the most popular topics) topic (page #s)

1. Getting started (3-6)
2. "Welcome to Manatee" page and links (7-11,21,23,26-28)
3. "Genome Summary" page and links (11-20)
  - Annotation Notebook (15,37)
  - Genome Calculations (13)
  - Role Category Breakdown (14)
4. "Annotation Tools" page and links (28-38)
  - Gene List (34-38)
  - coordinate range (29)
  - overlaps (30)
  - InterEvidence (31)
5. Gene Curation Page (39-86)
  - BER section (43-47)
  - HMM section (55-57)
  - GO section (71-75,81)
6. Gene Ontology (21-22,71-81)
  - edit Gene Ontology (22)
  - search Gene Ontology (22,76-80)
  - Gene Ontology on the Gene Curation Page (71-75,81)
7. Genome Properties (23-25,57-60)
8. Genome Viewer (26,87-91)
9. TIGR role categories (35-36,38,82)
  - Role notes (38)
  - TIGR role entry on Gene Curation Page (82)
10. Edit starts (90)
11. Annotation Checklist (92)

# What Manatee Is

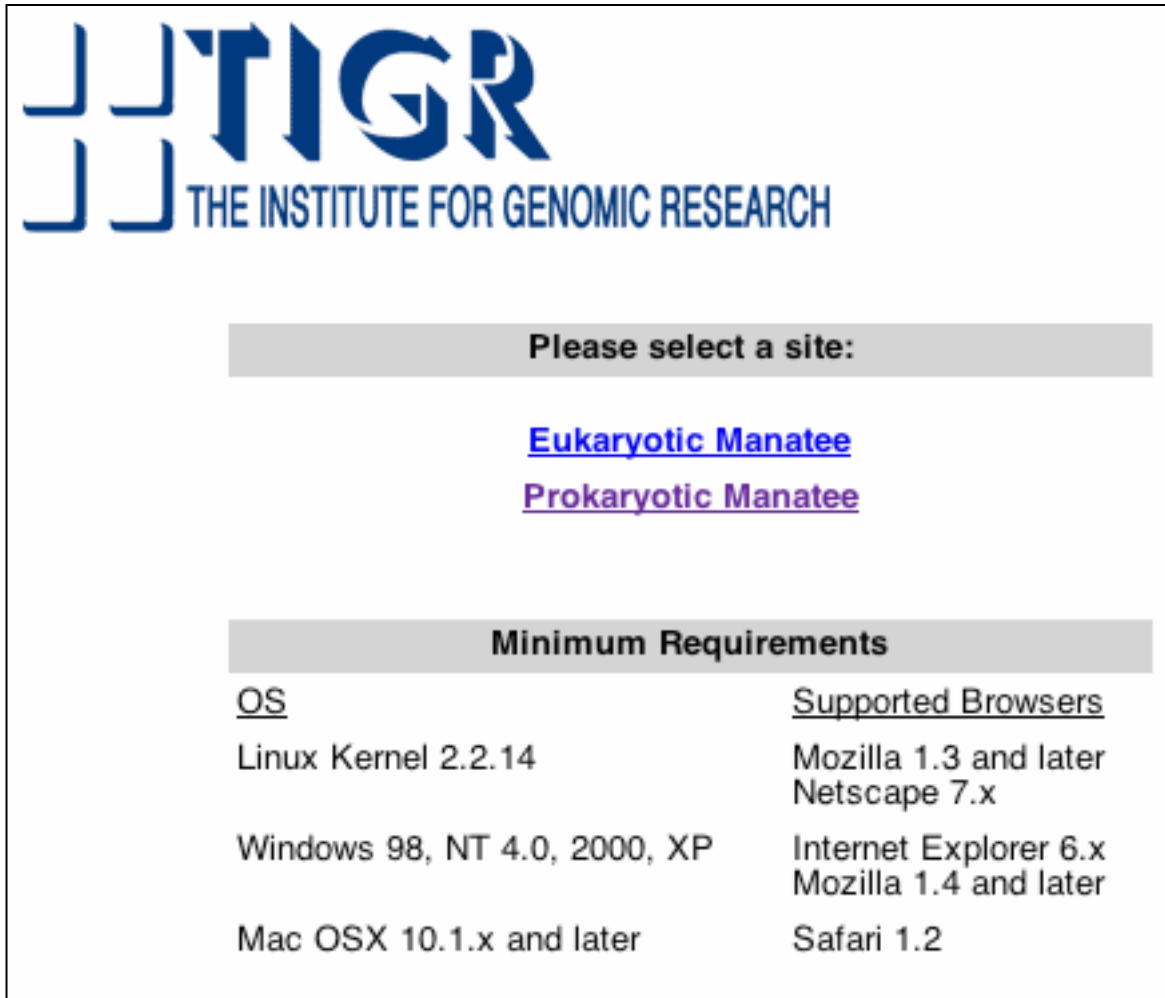
- Manatee is a web-based manual annotation tool for accessing and editing annotation data
- Manatee draws information from an underlying database for its displays
- Manatee sends information entered by annotators to the underlying database for storage
- Manatee depends on TIGR's database structure (more on this later)
- Multiple users can access the same database from different computers when Manatee is run on a server (as it is at TIGR)
- Installation questions will be addressed tomorrow by Todd Creasy, our Manatee developer (for those attending the 3-day course)

# Getting started with Manatee

- To log into Manatee within TIGR one must have a Sybase account and password.
- For those taking the 3-day course we have established a “training” account. This account is:
  - username = “training”
  - password = “training”
- TIGR employees have their own Sybase accounts and passwords.
- When logging into Manatee, one must enter a user name, a password, and the name of the database on which you wish to work.
- TIGR database names tend to be 3-5 letter codes:
  - during this tutorial and subsequent exercises we will be using the *Shewanella oneidensis* (formerly *Shewanella putrefaciens*) database.
  - we will be working with two versions of the *Shewanella* database:
    - the production database, which stores the published annotation (gsp)
    - a copy of the production database that has been set back to an un-annotated state (tgsp)

# Finding Manatee (working at TIGR)

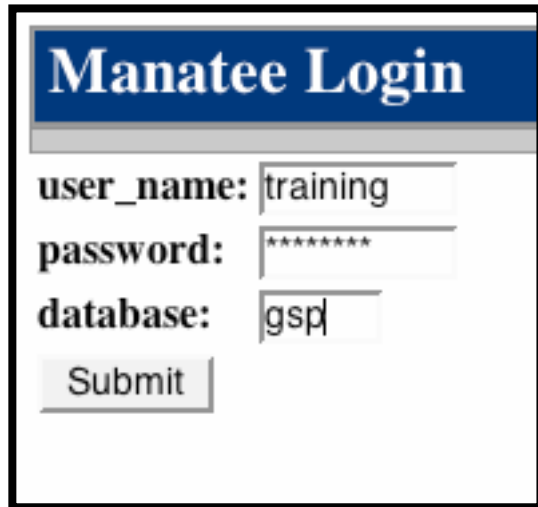
GO to [manatee.tigr.org](http://manatee.tigr.org) and select “Prokaryotic Manatee”



The screenshot shows the TIGR (The Institute for Genomic Research) website. At the top is the TIGR logo. Below it is a grey bar with the text "Please select a site:". Underneath this bar are two links: "Eukaryotic Manatee" in blue and "Prokaryotic Manatee" in purple. Below these links is another grey bar with the text "Minimum Requirements". Underneath this bar is a table with two columns: "OS" and "Supported Browsers".

| <u>OS</u>                    | <u>Supported Browsers</u>                      |
|------------------------------|------------------------------------------------|
| Linux Kernel 2.2.14          | Mozilla 1.3 and later<br>Netscape 7.x          |
| Windows 98, NT 4.0, 2000, XP | Internet Explorer 6.x<br>Mozilla 1.4 and later |
| Mac OSX 10.1.x and later     | Safari 1.2                                     |

Clicking on “Prokaryotic Manatee” from [www.manatee.org](http://www.manatee.org) takes you to the **Manatee Login Page**

A screenshot of a web form titled "Manatee Login". The form has a blue header with the title in white. Below the header, there are three input fields: "user\_name:" with the value "training", "password:" with the value "\*\*\*\*\*", and "database:" with the value "gsp". A "Submit" button is located at the bottom left of the form.

**Manatee Login**

user\_name: training

password: \*\*\*\*\*

database: gsp

Submit

Fill in the fields with the required information.

**user name and password:**

**3-day course:** user name and password are both “training”

**TIGR employees:** fill in your values

**database** = “gsp” for the tutorial portion and “tgsp” for the exercise portion (we are now in the tutorial portion).

# “Welcome to Manatee”

After logging in to Manatee, you come to the “Welcome to Manatee” page.

Here you will find several menu options and a couple search options to choose from.

I will discuss each in more detail in following slides.

NOTE: in the upper right hand corner of every Manatee page will be something like this:

[Home](#) | [Logged into \[gsp\] as mlgwinn](#)

The “Home” link takes you back to the “Welcome to Manatee” page, from where ever you are within the Manatee tool.

This area also shows you which database you are logged into, and who is logged in. Clicking on the login name will take you back to the login page.

Welcome to Manatee

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...

**ACCESS LISTINGS**

- ▶ Annotation Tools
- ▶ Genome Summary
- ▶ Gene Ontology
- ▶ Genome Properties
- ▶ Genome Viewer
- ▶ Multi Genome Annotation Tool

submit

reset

**ACCESS GENE CURATION PAGE**

▶ gene:

**SEARCH GENES BY GENE NAME**

▶ gene name:

**CHANGE ORGANISM DATABASE**

▶ database:

# The Welcome to Manatee Page

## “Access Gene Curation Page” option

We will look at the options in the Access Listings section in subsequent slides. First we will look at the 3 options on the bottom of this page:

### Access Gene Curation Page:

This option will take you directly to a page containing gene specific information called the “Gene Curation Page” or “GCP” for short. The GCP displays most of what knowledge we have about a given protein - you will be seeing this page in much more detail later. For now just know that you can reach this page by entering either a feat\_name or locus id into this box and then clicking “submit”. A feat\_name is an internal identifier given to each gene in a genome, feat\_names are not used publically. These are initially assigned by Glimmer and generally are numbered sequentially from the beginning of the DNA sequence given to Glimmer. They have the format ORF#####, where ORF stands for “open reading frame” and ##### is a 5-digit zero padded number. (For more on this see the overview document.) Locus ids (loci) are assigned to proteins at the end of the annotation process. They are numbered sequentially from the origin of replication of the genome (if it can be identified). Loci are unique accessions and are used for public release and display of the proteins.

**Welcome to Manatee**

 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 s

**ACCESS LISTINGS**

- ▶ Annotation Tools
- ▶ Genome Summary
- ▶ Gene Ontology
- ▶ Genome Properties
- ▶ Genome Viewer
- ▶ Multi Genome Annotation Tool

☐ **ACCESS GENE CURATION PAGE**

- ▶ gene:

☐ **SEARCH GENES BY GENE NAME**

- ▶ gene name:

☐ **CHANGE ORGANISM DATABASE**

- ▶ database:

# The Welcome to Manatee Page

## “Search Genes By Gene Name” option

This is a keyword based search for the common names that have been given to the genes/proteins (we have a tendency to use the terms gene and protein somewhat interchangeably, however, what we are really annotating are the protein translations of the predicted genes.)

Whatever keyword you enter will be treated as though it has wildcards flanking it. This means that you will get results that include names containing your keyword as an individual word and names that contain words that contain your keyword.

For example, if you search with “kinase”

you could get these:

“adenylate kinase”

“protein kinase”

“sensor histidine kinase”

as well as these:

“glutamate 5-kinase”

“phosphoenolpyruvate carboxykinase”

“ribose-phosphate pyrophosphokinase”

The results will be in the form of a table containing additional information and links to other pages - this table format will be described later.

**Welcome to Manatee**

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 e

**ACCESS LISTINGS**

- ▶ **Annotation Tools**
- ▶ **Genome Summary**
- ▶ **Gene Ontology**
- ▶ **Genome Properties**
- ▶ **Genome Viewer**
- ▶ **Multi Genome Annotation Tool**

**ACCESS GENE CURATION PAGE**

▶ gene:

**SEARCH GENES BY GENE NAME**

▶ gene name:

**CHANGE ORGANISM DATABASE**

▶ database:

## The Welcome to Manatee Page “Change Organism Database” option

To change from one database to another, one does not need to re-login, rather one need only type in the name of the database they wish to go to and click submit.

**Welcome to Manatee**  
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 e

submit

reset

**ACCESS LISTINGS**

- ▶ Annotation Tools
- ▶ Genome Summary
- ▶ Gene Ontology
- ▶ Genome Properties
- ▶ Genome Viewer
- ▶ Multi Genome Annotation Tool

**ACCESS GENE CURATION PAGE**  
▶ gene:

**SEARCH GENES BY GENE NAME**  
▶ gene name:

**CHANGE ORGANISM DATABASE**  
▶ database:

## The Welcome to Manatee Page Options under “Access Listings”: “Genome Summary”

The “Annotation Tools” option is one of the most used and will be described in detail in later slides. The “Gene Ontology”, “Genome Properties”, and “Genome Viewer” sections are accessible here as well as elsewhere within Manatee. (There are many routes to view the various pieces of information within Manatee.) They will be described briefly as links from “Access Listings” and then in more detail as they are viewed from the Gene Curation Page (GCP) and elsewhere.

First we will look in more detail at the options under “**Genome Summary**”

Clicking on “Genome Summary” takes one to a new page with additional menu options (on next slide).

The screenshot shows the 'Welcome to Manatee' page. At the top is a blue header with the text 'Welcome to Manatee'. Below it is a light gray box containing a document icon and the text: '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 e'. The main content area has a light blue background. On the left side, there are two buttons: 'submit' and 'reset'. The main content area is divided into three sections, each with a light blue header bar. The first section is 'ACCESS LISTINGS', which contains a list of links: 'Annotation Tools', 'Genome Summary' (highlighted with a red box), 'Gene Ontology', 'Genome Properties', 'Genome Viewer', and 'Multi Genome Annotation Tool'. The second section is 'ACCESS GENE CURATION PAGE', which contains a 'gene:' label and a text input field. The third section is 'SEARCH GENES BY GENE NAME', which contains a 'gene name:' label and a text input field. The fourth section is 'CHANGE ORGANISM DATABASE', which contains a 'database:' label and a text input field.

Welcome to Manatee

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 e

**ACCESS LISTINGS**

- ▶ Annotation Tools
- ▶ **Genome Summary**
- ▶ Gene Ontology
- ▶ Genome Properties
- ▶ Genome Viewer
- ▶ Multi Genome Annotation Tool

submit

reset

**ACCESS GENE CURATION PAGE**

▶ gene:

**SEARCH GENES BY GENE NAME**

▶ gene name:

**CHANGE ORGANISM DATABASE**

▶ database:

# The “Genome Summary” page

**Genome Summary Page**[Home](#) | [Logged into \[gsi\] as mlgwinn](#)

The Genome Summary information page displays specific information concerning the selected genome. From here the user can view the following information : ORF counts, Role category information, genes of interest, HMM and paralogous family information, membrane protein information, frameshift information, and annotation progress.

[Home](#)[Annotation Tools](#)[Genome Summary](#)[Gene Assignment Help](#)

**SUMMARY LISTS**

- ▶ **Genome Calculations**
- ▶ **Role Category Breakdown**
- ▶ **Annotation Notebook**
- ▶ **Project Administration**
- ▶ **Frameshift Status**
- ▶ **Annotation Progress Report**
- ▶ **By Interpro Domain**
- ▶ **Genome Properties**

**ATTRIBUTES**

▶ **attribute:**  **order by:**

**EVIDENCE**

▶ **evidence:**  **order by:**

**PARALOUS PROTEINS**

▶ **view by:**

**MEMBRANE PROTEINS**

▶ **sort by:**

▶ **signal peptide cutoffs:**  
 **Y-score lower limit**  
 **S-mean lower limit**

▶ **transmembrane regions:**  
 **lower limit**  
 **upper limit**

▶ **constraints:**  
☐ **genes with OMP signals**  
☐ **genes with lipid attachment sites**

Clicking on the item in the list of options takes you to a page with the information or more options. Following slides will describe each of these.

These are tools that allow one to view the data based on various types of annotation. Following slides will describe the use and output of each of these.

# Links from the Genome Summary Page: “Genome Calculations”

| feature name                                 | feature count | feature type |
|----------------------------------------------|---------------|--------------|
| ▶ <a href="#">Open Reading Frame</a>         | 4930          | ORF          |
| ▶ <a href="#">ribosome binding site</a>      | 4444          | RBS          |
| ▶ <a href="#">rho-independent terminator</a> | 846           | TERM         |
| ▶ <a href="#">Reject ORFS</a>                | 169           | RORF         |
| ▶ <a href="#">transfer RNA</a>               | 101           | tRNA         |
| ▶ <a href="#">ribosomal RNA</a>              | 27            | rRNA         |
| ▶ <a href="#">structural RNA</a>             | 3             | sRNA         |
| ▶ <a href="#">Bacteriophage</a>              | 3             | PHAGE        |

| start sites              | number         | percent          |
|--------------------------|----------------|------------------|
| ▶ <a href="#">ATG:</a>   | 3933<br>(3235) | 79.8%<br>(85.9%) |
| ▶ <a href="#">GTG:</a>   | 541<br>(350)   | 11.0%<br>(9.3%)  |
| ▶ <a href="#">TTG:</a>   | 453<br>(182)   | 9.2%<br>(4.8%)   |
| ▶ <a href="#">OTHER:</a> | 1<br>(1)       | 0.0%<br>(0.0%)   |

Numbers in parentheses do not include hypothetical proteins

This page shows the various calculable and countable features of the genome. This information is newly generated each time the page is accessed so that all information is current.

| 'chromosome' Information Table             |            |       |       |       |
|--------------------------------------------|------------|-------|-------|-------|
| ▶ sequence id:                             | 7974       |       |       |       |
| ▶ type:                                    | chromosome |       |       |       |
| ▶ molecule length:                         | 4969803 bp |       |       |       |
| ▶ GC content:                              | 46%        |       |       |       |
| ▶ base frequencies:                        | (A)        | (C)   | (G)   | (T)   |
|                                            | 27.0%      | 23.0% | 23.0% | 27.0% |
| ▶ funny characters:                        | R          | Y     |       |       |
|                                            | 2          | 6     |       |       |
| ▶ ORF count:                               | 4758       |       |       |       |
| ▶ average gene length:                     | 895 nt     |       |       |       |
| ▶ percent coding:                          | 85.8%      |       |       |       |
| ▶ percent coding OR tRNA, rRNA, or repeat: | 90.0%      |       |       |       |
| ▶ skew table                               |            |       |       |       |

## Links from the Genome Summary Page: “Role Category Breakdown”

This page shows a summary of the genes found in various broad categories based on TIGR roles and then a breakdown by TIGR sub category. Each blue role id number or “main” is a link to a table containing a list of all the genes in that category.

| ► ORF Summary          |             |                |
|------------------------|-------------|----------------|
| <b>Total ORFs:</b>     | <b>4930</b> | <b>100.0 %</b> |
| assigned function      | 2521        | 51.1 %         |
| conserved hypothetical | 871         | 17.7 %         |
| unknown function       | 378         | 7.7 %          |
| hypothetical proteins  | 1162        | 23.6 %         |

| ► Role Breakdown     |                                                    |        |          |       |
|----------------------|----------------------------------------------------|--------|----------|-------|
| role id              | name                                               | number | complete | %     |
| <a href="#">main</a> | Unclassified                                       | 2      | 0        | 0.04% |
| <a href="#">185</a>  | Role category not yet assigned                     | 2      | 0        | 0.04% |
| <a href="#">main</a> | Amino acid biosynthesis                            | 91     | 0        | 1.85% |
| <a href="#">70</a>   | Aromatic amino acid family                         | 17     | 0        | 0.34% |
| <a href="#">71</a>   | Aspartate family                                   | 24     | 0        | 0.49% |
| <a href="#">73</a>   | Glutamate family                                   | 21     | 0        | 0.43% |
| <a href="#">74</a>   | Pyruvate family                                    | 13     | 0        | 0.26% |
| <a href="#">75</a>   | Serine family                                      | 8      | 0        | 0.16% |
| <a href="#">161</a>  | Histidine family                                   | 8      | 0        | 0.16% |
| <a href="#">69</a>   | Other                                              | 0      | 0        | 0.00% |
| <a href="#">main</a> | Purines, pyrimidines, nucleosides, and nucleotides | 63     | 0        | 1.28% |
| <a href="#">123</a>  | 2'-Deoxyribonucleotide metabolism                  | 8      | 0        | 0.16% |
| <a href="#">124</a>  | Nucleotide and nucleoside interconversions         | 11     | 0        | 0.22% |

Links from the Genome Summary Page: “Annotation Notebook” - the annotation notebook is a set of text fields associated with each TIGR role category. These are used for annotators to store information about the annotation which they feel the PIs of the project should know for purposes of writing the manuscript, generally they consist of items of particular biological interest, often involving the presence or absence of particular pathways, genes, gene order, etc. These entries are entered and edited with the “Edit Annotation Notebook” page, linked from the gene list, see page 33, 36 of this tutorial.

|                                                                                                                                                                                                                                                                                                               |                            |                                              |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|----------------------------------------------|
| <b><i>Shewanella oneidensis</i> MR-1</b>                                                                                                                                                                                                                                                                      | <b>Annotation Notebook</b> | Logged into [gsp] as <a href="#">mlgwinn</a> |
| <p> The annotation notebook is used by annotators to note genes of interest for the project leader. The notebook is organized by role category. The project leader can view the page, and view the genes at their leisure.</p> |                            |                                              |

|                                                                                                                                                                                                                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Biosynthesis of cofactors, prosthetic groups, and carriers</b>                                                                                                                                                                                                                 |
| <b>83 : Pantothenate and coenzyme A</b>                                                                                                                                                                                                                                           |
| Appears to lack panD which I searched for with the E.coli sequence. No matches using blastp or tblastn. RTD                                                                                                                                                                       |
| <b>84 : Pyridoxine</b>                                                                                                                                                                                                                                                            |
| Appears to lack serC which I searched for with the Caulobacter sequence. No high-scoring matches using blastp or tblastn. RTD                                                                                                                                                     |
| <b>Cellular processes</b>                                                                                                                                                                                                                                                         |
| <b>96 : Detoxification</b>                                                                                                                                                                                                                                                        |
| arsenate reductase operon quirky. Sequences diverge from typical prokaryotic arsB and arsC. Order of these two genes in the operon is reversed as well. The arsenate reductase repressor arsR is well conserved. There may also be a stand alone version of arsC present as well. |
| <b>Protein fate</b>                                                                                                                                                                                                                                                               |
| <b>97 : Protein and peptide secretion and trafficking</b>                                                                                                                                                                                                                         |
| general secretion pathway genes are split into two operons (gspAB) and (gspCDEFGHIJKLMN). Not sure how typical this is.                                                                                                                                                           |
| <b>Biosynthesis of cofactors, prosthetic groups, and carriers</b>                                                                                                                                                                                                                 |
| <b>162 : Thiamine</b>                                                                                                                                                                                                                                                             |
| Appear to lack some components of the the pathway for thiamine biosynthesis. RTD                                                                                                                                                                                                  |

## Other links from the Genome Summary Page:

**“Project Administration”** - Clicking this link takes one to a page that displays the administrative information for the project: things like PI, grant #, etc.

**“Frameshift Status”** - Currently this tool is not available for people running Manatee locally outside of TIGR. For TIGR users, there is (will be) a separate section of the tutorial governing all things involving Frameshifts (this part of our SOP is currently undergoing change.) In brief, this link displays a page listing all of the genes in the genome which needed to be reviewed for the presence of a frameshift or in-frame stop codon as well as the status of each.

**“Annotation Progress Report”** - This links to a page that lists all of the processes that must be carried out during the annotation of a genome and provides fields in which to enter when each process was done and who did it. There is also a link to a page listing all the TIGR mainrole categories and fields for individual annotators to sign up for each category.

**“by InterPro Domain”** - This links to a list of genes according to membership in an InterPro domain.

**“Genome Properties”** - another link to this tool set, will be described in detail elsewhere.

# Searches on the Genome Summary page: “Attributes”

## ATTRIBUTES

attribute:  order by:

One can choose to view genes based on one of several “attributes” they might have. Here I have shown a selection for “MW” which stands for molecular weight. Once you choose an attribute to search by, you can then choose various ordering display options. The above choices will show the proteins in the genome according to calculated molecular weight with the heaviest ones first. (see below)

| <i>Silicibacter pomeroyi</i> DSS-3 |                                 | Attribut |
|------------------------------------|---------------------------------|----------|
|                                    |                                 |          |
| PAGE 1 2 3 4 5                     |                                 |          |
| MW                                 | <input type="text" value="MW"/> |          |
| gene id                            | score                           |          |
| <a href="#">ORF03439</a>           | 831333.58                       |          |
| <a href="#">ORF01043</a>           | 231337.37                       |          |
| <a href="#">ORF04635</a>           | 225004.71                       |          |
| <a href="#">ORF00157</a>           | 192587.87                       |          |
| <a href="#">ORF03448</a>           | 170927.61                       |          |
| <a href="#">ORF00679</a>           | 170039.51                       |          |

This is just the top of a very long list containing all of the proteins in the genome. One can click on any of the blue gene id links and get go to the Gene Curation Page (GCP) for the gene. (The GCP will be described in detail shortly.)

One can jump to different pages in the list by clicking on the blue numbers in the boxes above the list.

One can change the order of the list by clicking in the arrows in blue circles.

# Searches on the Genome Summary page: “Evidence”

## EVIDENCE

► evidence:  order by:

One can choose to view the genes based on one of several types of clustering evidence that has been found for them, some as the result of InterPro searches and some as a result of separate searches we perform. Here, I have selected “HMM2” (which will include both the TIGRFAM and Pfam HMM sets) and I will view the output ordered by the number of hits in the genome, the HMMs with the most hits will be listed first.

| Shewanella oneidensis MR-1 Evidence Display |           |                                                                 |   |   |   |   |   |   |   |    |    |    |
|---------------------------------------------|-----------|-----------------------------------------------------------------|---|---|---|---|---|---|---|----|----|----|
| PAGE                                        | 1         | 2                                                               | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| HMM2                                        |           |                                                                 |   |   |   |   |   |   |   |    |    |    |
| count                                       | accession | HMM name                                                        |   |   |   |   |   |   |   |    |    |    |
| 85                                          | PF00005   | ABC transporter                                                 |   |   |   |   |   |   |   |    |    |    |
| 82                                          | TIGR01199 | Helix-turn-helix domain, fis-type                               |   |   |   |   |   |   |   |    |    |    |
| 71                                          | PF00072   | Response regulator receiver domain                              |   |   |   |   |   |   |   |    |    |    |
| 69                                          | PF02653   | Branched-chain amino acid transport system / permease component |   |   |   |   |   |   |   |    |    |    |
| 62                                          | PF00665   | Integrase core domain                                           |   |   |   |   |   |   |   |    |    |    |
| 62                                          | PF00672   | HAMP domain                                                     |   |   |   |   |   |   |   |    |    |    |
| 59                                          | PF00037   | 4Fe-4S binding domain                                           |   |   |   |   |   |   |   |    |    |    |
| 57                                          | PF00271   | Helicase conserved C-terminal domain                            |   |   |   |   |   |   |   |    |    |    |
| 56                                          | PF00872   | transposase, Mutator family                                     |   |   |   |   |   |   |   |    |    |    |
| 56                                          | PF02796   | Helix-turn-helix domain of resolvase                            |   |   |   |   |   |   |   |    |    |    |
| 54                                          | PF03466   | LysR substrate binding domain                                   |   |   |   |   |   |   |   |    |    |    |
| 53                                          | PF00486   | Transcriptional regulatory protein, C terminal                  |   |   |   |   |   |   |   |    |    |    |
| 53                                          | PF00990   | GGDEF domain                                                    |   |   |   |   |   |   |   |    |    |    |

By clicking on the blue numbers in each row, one will get a list of genes that hit that HMM. Clicking on the blue accession number will take one to an info page for the HMM in question.

One can reorder the list by count or accession by clicking on the blue column headers.

Numbered boxes at the top will take one to a desired page in the output.

# Searches on the Genome Summary page: “Paralogous Families”

## PARALOGOUS PROTEINS

view by:

One can choose to view the genes based on membership in paralogous families, ordering either by number of family members or by family name.

### *Silicibacter pomeroyi* DSS-3 Paralogous Families

The Paralogs page comes in two flavors. If launched from the Gene Summary page, it displays a summary (paralog page is launched from itself (i.e. from the paralog skim flavor) it displays an ORF list of all members for the selected pa

total families: 762 total proteins in families: 2647

| proteins | family name   | description                                                             |
|----------|---------------|-------------------------------------------------------------------------|
| 114      | fam_PF00005   | ABC transporter                                                         |
| 82       | fam_PF00528   | Binding-protein-dependent transport systems inner membrane component    |
| 67       | fam_PF00126   | transcriptional regulator, LysR family                                  |
| 60       | fam_PF03466   | LysR substrate binding domain                                           |
| 44       | fam_TIGR01409 | Tat (twin-arginine translocation) pathway signal sequence               |
| 37       | fam_PF00106   | oxidoreductase, short chain dehydrogenase/reductase family              |
| 34       | fam_PF00072   | Response regulator/receiver domain                                      |
| 33       | fam_PF00892   | Integral membrane protein DUF6                                          |
| 31       | fam_PF02653   | Branched-chain amino acid transport system / permease component         |
| 30       | fam_PF00392   | transcriptional regulator, gntR family                                  |
| 29       | fam_PF00583   | acetyltransferase, GNAT family                                          |
| 27       | fam_PF02518   | ATPase, histidine kinase-, DNA gyrase B-, and HSP90-like domain protein |
| 25       | fam_PF06808   | TRAP transporter, DctM-like membrane protein                            |
| 25       | fam_PF00165   | transcriptional regulator, AraC family                                  |

-Paralogous families are built by first searching all of the proteins within a genome against themselves and against the HMM db. If a paralogous family matches an HMM the family will be named based on the HMM. Then further searches are done to group the proteins based on regions of sequence that did not match an HMM. Those families are given numerical names and do not have descriptions.

-Output shows you the number of members in each family, the name of the family, and a description of the family (if the family is based on an HMM).

-You can view a list of the proteins in each family by clicking on the family name. You can view information about the HMM on which the family is based by clicking on the description.

# Searches on the Genome Summary page: “Membrane proteins”

## MEMBRANE PROTEINS

sort by:

signal peptide cutoffs:

Y-score lower limit

S-mean lower limit

transmembrane regions:

lower limit

upper limit

constraints:

☐ genes with OMP signals

☐ genes with lipid attachment sites

One can choose to view the proteins based on predicted location in a membrane. You can choose particular SignalP cutoff values, number of predicted transmembrane regions, proteins that have an OMP signal, or lipid attachment site. You can also sort the output by several different options.

| Silicibacter pomeroyi DSS-3 Membrane Proteins                                                                                                                                                                                                                                                                                                 |                                      |                           |      |             |     |    | <a href="#">Home</a>   <a href="#">Logged into [gsi] as mlgwinn</a>  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|---------------------------|------|-------------|-----|----|----------------------------------------------------------------------|
| This page displays an ORF list based on the results of four membrane protein analyses: signal peptide sequence, membrane spanning regions, outer membrane protein sequence, and lipoprotein sequence. The user controls the output by specifying one to four of these criteria in any combination as well as the sorting key for the results. |                                      |                           |      |             |     |    |                                                                      |
| sort by:                                                                                                                                                                                                                                                                                                                                      | <input type="text" value="gene id"/> | Sort                      |      |             |     |    |                                                                      |
| gene id                                                                                                                                                                                                                                                                                                                                       | y-score<br>( $\geq 0.36$ )           | s-mean<br>( $\geq 0.54$ ) | site | TM<br>(3/5) | OMP | LP | gene name                                                            |
| ORF00005                                                                                                                                                                                                                                                                                                                                      | 0.372                                | 0.611                     | 91   | 3           |     |    | conserved hypothetical protein                                       |
| ORF00119                                                                                                                                                                                                                                                                                                                                      | 0.557                                | 0.739                     | 31   | 4           |     |    | peptide/opine/nickel uptake family ABC transporter, permease protein |
| ORF00160                                                                                                                                                                                                                                                                                                                                      | 0.546                                | 0.723                     | 53   | 3           |     |    | conserved hypothetical protein                                       |
| ORF00234                                                                                                                                                                                                                                                                                                                                      | 0.692                                | 0.835                     | 25   | 5           |     |    | flagellar biosynthetic protein FlIP                                  |
| ORF00416                                                                                                                                                                                                                                                                                                                                      | 0.407                                | 0.825                     | 15   | 5           |     |    | NnrU family protein                                                  |
| ORF00422                                                                                                                                                                                                                                                                                                                                      | 0.387                                | 0.568                     | 49   | 3           |     |    | succinate dehydrogenase, hydrophobic membrane anchor protein         |
| ORF00522                                                                                                                                                                                                                                                                                                                                      | 0.411                                | 0.613                     | 24   | 3           |     |    | hypothetical protein                                                 |
| ORF00617                                                                                                                                                                                                                                                                                                                                      | 0.478                                | 0.628                     | 22   | 4           |     |    | sterol desaturase, homolog                                           |
| ORF00725                                                                                                                                                                                                                                                                                                                                      | 0.434                                | 0.732                     | 28   | 3           |     |    | conserved hypothetical protein                                       |
| ORF01154                                                                                                                                                                                                                                                                                                                                      | 0.687                                | 0.576                     | 90   | 5           |     |    | membrane protein, putative                                           |
| ORF01297                                                                                                                                                                                                                                                                                                                                      | 0.434                                | 0.544                     | 35   | 4           |     |    | membrane protein, putative                                           |
| ORF01321                                                                                                                                                                                                                                                                                                                                      | 0.654                                | 0.724                     | 33   | 4           |     |    | hypothetical protein                                                 |
| ORF01507                                                                                                                                                                                                                                                                                                                                      | 0.396                                | 0.585                     | 41   | 5           |     |    | oligopeptide ABC transporter, permease protein                       |
| ORF01581                                                                                                                                                                                                                                                                                                                                      | 0.463                                | 0.610                     | 29   | 3           |     |    | hypothetical protein                                                 |
| ORF01589                                                                                                                                                                                                                                                                                                                                      | 0.370                                | 0.547                     | 29   | 3           |     |    | conserved hypothetical protein                                       |
| ORF01645                                                                                                                                                                                                                                                                                                                                      | 0.565                                | 0.688                     | 57   | 4           |     |    | lipoprotein, putative                                                |

Output shows a table of the genes with the chosen parameters. You can reorder them using the pull-down menu and the “sort” button. The table displays all of the parameters available for each protein. Clicking on the blue gene id takes you to the Gene Curation Page (GCP) for the gene.

# The Welcome to Manatee Page

## Options under “Access Listings”: “Gene Ontology”

This link will open a page that offers options for using the Gene Ontology (GO) system.

(For more information on the Gene Ontology system, see the Annotation Overview document, or the Gene Ontology web site, [www.geneontology.org](http://www.geneontology.org))

In brief, the GO offers a controlled vocabulary for the description of aspects of gene products. Currently, TIGR assigns both TIGR role categories and GO terms to all of our genes. Manatee has many built in features for the suggestion and entry of GO terms and associated information. These features will be detailed in later slides. The next slide shows a brief description of the links available here and of the “edit GO” options.

When Manatee refers to “editing” GO, we mean the creation of “TI” or TIGR terms. These are temporary terms created for use in-house at TIGR until corresponding terms are created at GO. When a need for a new term is found, we (usually Michelle) submit a request to the GO via their SourceForge tracking site that the new term be created. If a TIGR annotator needs the new term right away, they can create a TI term to use within our db. Later, when the official GO term is made, the TI term id will be replaced with the new GO term id.

Welcome to Manatee

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...

### ACCESS LISTINGS

- ▶ Annotation Tools
- ▶ Genome Summary
- ▶ Gene Ontology
- ▶ Genome Properties
- ▶ Genome Viewer
- ▶ Multi Genome Annotation Tool

submit

### ACCESS GENE CURATION PAGE

▶ gene:

reset

### SEARCH GENES BY GENE NAME

▶ gene name:

### CHANGE ORGANISM DATABASE

▶ database:

# The Welcome to Manatee page, links from Access Listings: "Gene Ontology"

**Gene Ontology Page**

Choose to **search** or **edit** GO.

**ACCESS GENE ONTOLOGY**

- ▶ **Search Gene Ontology**
- ▶ **Edit Gene Ontology**

When we refer to editing the GO in TIGR's db, we are referring the creation of TI terms. (see overview for more information)

**Ontology Edit Page**

**EDIT ONTOLOGY**

- ▶ **Add Ontology Id**
- ▶ **Create Ontology Link**
- ▶ **Update Ontology Term**
- ▶ **View TI Status**

These are links to pages that allow you to enter, update, and add parents to a TI term.

**Add Ontology Page**

add new ontology id:

|                    |                      |
|--------------------|----------------------|
| name               | <input type="text"/> |
| definition         | <input type="text"/> |
| parent ontology id | <input type="text"/> |

☐ isa ☐ partof

search GO from GO id: search GO from EC number: search GO from keywords:

GO id:  EC #:

☒ Exact Match ☐ Case Sensitive

search GO associations GO id:

search GO correlations:

☐ Search TIGR prokaryotic data only

Search options will be discussed in detail later

This links to a page that displays all TI terms and their status.

**TI Status Page** [Home](#) | [Logged into \[gsp\] as mlgwin](#)

| TI term                              | GO id replacement | name                                          | type               | assigned by | last update |
|--------------------------------------|-------------------|-----------------------------------------------|--------------------|-------------|-------------|
| TI:0000001 <a href="#">[replace]</a> | none              | Ontology                                      | root               | bhaas       | 02/11/02    |
| TI:0000010                           | GO:0012505        | endomembrane system                           | cellular_component | lhannick    | 02/22/02    |
| TI:0000013                           | GO:0005770        | prevacuolar compartment (sensu Viridiplantae) | cellular_component | lhannick    | 02/27/02    |
| TI:0000015                           | GO:0042262        | DNA protection                                | biological_process | access      | 03/11/02    |
| TI:0000016                           | GO:0008738        | L-fucose phosphate aldolase                   | molecular_function | access      | 03/14/02    |
| TI:0000018                           | GO:0045152        | anti sigma factor antagonist                  | molecular_function | access      | 03/19/02    |
| TI:0000019                           | GO:0042243        | spore wall assembly (sensu Bacteria)          | biological_process | access      | 03/20/02    |
| TI:0000021                           | GO:0045148        | tripeptide aminopeptidase                     | molecular_function | access      | 03/22/02    |
| TI:0000022                           | GO:0045149        | acetoin metabolism                            | biological_process | mlgwin      | 03/27/02    |
| TI:0000023                           | GO:0045151        | acetoin biosynthesis                          | biological_process | mlgwin      | 03/27/02    |

## The Welcome to Manatee Page Options under “Access Listings”: “Genome Properties”

The Genome Properties system allows one to view annotation from the context of the whole genome. It predicts and/or captures information on the presence/absence of pathways, cellular structures and other features of the organism. (see overview for more details) Clicking on the Genome Properties link from the “Welcome to Manatee” page displays a table of all of the properties and their states for the organism you are working on. The state is “yes” if the property is present, “no” if the property is absent, and will have other intermediate values such as “some evidence” or “not supported” depending on the amount of evidence for a given property. Details on what is known about each property for the genome you are working on can be obtained by clicking on the blue property name. (see next slide)

Welcome to Manatee

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 e

**ACCESS LISTINGS**

- ▶ Annotation Tools
- ▶ Genome Summary
- ▶ Gene Ontology
- ▶ **Genome Properties**
- ▶ Genome Viewer
- ▶ Multi Genome Annotation Tool

submit

reset

**ACCESS GENE CURATION PAGE**

▶ gene:

**SEARCH GENES BY GENE NAME**

▶ gene name:

**CHANGE ORGANISM DATABASE**

▶ database:

## The Welcome to Manatee page, links from Access Listings: “Genome Properties”

update genome properties

|                                     |            |
|-------------------------------------|------------|
| property                            | state      |
| 2-aminoethylphosphonate degradation | NULL       |
| curator comment                     | prediction |
|                                     | NULL       |
| assign by                           |            |
| submit                              |            |

Update the status of or information about a property in this organism

search genome properties

Search By: Name  submit reset

Search for a property in this genome.

**Shewanella oneidensis MR-1** **Genome Properties** [Home](#) | Logged into [gsp] as [mlgwin](#)

[Update Property](#) [Search Property](#)

| property                                                                     | state         | value | assign by | prediction | experiment | curator comment | auto comment                             |
|------------------------------------------------------------------------------|---------------|-------|-----------|------------|------------|-----------------|------------------------------------------|
| <a href="#">2-aminoethylphosphonate degradation</a>                          | none found    | 0     | RULES     | 1          |            |                 |                                          |
| <a href="#">4-hydroxyphenylacetate degradation</a>                           | not supported | 0.1   | RULES     | 1          |            |                 |                                          |
| <a href="#">5-aminolevulinate biosynthesis</a>                               | YES           |       | RULES     | 1          |            |                 |                                          |
| <a href="#">5-aminolevulinate biosynthesis (glutamate pathway)</a>           | YES           | 1     | HYBRID    | 1          |            |                 |                                          |
| <a href="#">5-aminolevulinate biosynthesis from succinyl-CoA and glycine</a> | NO            |       | RULES     | 1          |            |                 | none found AND glutamate pathway present |
| <a href="#">acetyl-CoA carboxylase complex</a>                               | some evidence | 0.25  | RULES     | 1          |            |                 |                                          |

Click on the blue name of a property to learn more about the steps/requirements for the property and to see background information and references regarding the property. You can also see the genes the are involved in the property in the context of their neighbors in the genome. These pages will be shown in detail later in the tutorial but are quickly shown on the next slide.

## Genome Property information Page (in brief, more detail will be shown later in the tutorial)

| Property Definition |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |           |                    |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------------------|
| property:           | arginine biosynthesis from ornithine, carbamoyl-p and aspartate                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | state:    | YES                |
| property type:      | PATHWAY                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | value:    | 1                  |
| role id:            | 73 [ <a href="#">update</a> ]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | assignby: | RULES              |
| GO ids:             | GO:0042450 [ <a href="#">view</a>   <a href="#">add</a>   <a href="#">update</a> ] arginine biosynthesis via ornithine                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | date:     | Apr 8 2003 11:09AM |
| description:        | <p>The arginine biosynthesis pathway is a three step process and a part of the urea cycle. [2] The first enzyme, ornithine carbamoyltransferase (OTCase) carries out the reaction anabolically in arginine biosynthesis, but in some cases, it carries out catabolic reactions. Most OTCases are homotrimers, but the homotrimers are organized into dodecamers built from four trimers in at least two species; the catabolic OTCase of <i>Pseudomonas aeruginosa</i> is allosterically regulated, while OTCase of the extreme thermophile <i>Pyrococcus furiosus</i> shows both allostery and thermophily. [1] The third step of the pathway yields L-arginine by argininosuccinate lyase, and the amino acid can be cleaved by arginase, yielding urea and reconstituting ornithine. L-arginine also can be utilized in the creatine biosynthesis.</p> |           |                    |

Information on the property.

| Property Steps                                                                     |                                                                                     |                                    |          |                                |           |         |             |           |    |
|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|------------------------------------|----------|--------------------------------|-----------|---------|-------------|-----------|----|
| REQUIRED                                                                           |                                                                                     | ornithine carbamoyltransferase (1) |          |                                |           |         |             |           |    |
| A                                                                                  | C                                                                                   | GC                                 | gene id  | gene name                      | evidence  | role id | gene symbol | EC number | OP |
|  |  |                                    | ORF02078 | ornithine carbamoyltransferase | TIGR00658 | 73      | argF        | 2.1.3.3   |    |
| REQUIRED                                                                           |                                                                                     | argininosuccinate synthase (2)     |          |                                |           |         |             |           |    |
| A                                                                                  | C                                                                                   | GC                                 | gene id  | gene name                      | evidence  | role id | gene symbol | EC number | OP |
|  |  |                                    | ORF02077 | argininosuccinate synthase     | TIGR00032 | 73      | argG        | 6.3.4.5   |    |
| REQUIRED                                                                           |                                                                                     | argininosuccinate lyase (3)        |          |                                |           |         |             |           |    |
| A                                                                                  | C                                                                                   | GC                                 | gene id  | gene name                      | evidence  | role id | gene symbol | EC number | OP |
|  |  |                                    | ORF02076 | argininosuccinate lyase        | TIGR00838 | 73      | argH        | 4.3.2.1   |    |

Information on the genes identified to be a part of the property.

The Welcome to Manatee page, links from Access Listings: “Genome Viewer”

Genome Viewer is a tool which allows one to view the genes in context with their neighboring genes in the genome. It displays a graphic showing the 6-frame translation of a region of DNA sequence, where each horizontal bar is a different frame. Arrows representing the genes are color coded according to TIGR mainrole assignment. There are many viewing and editing options available from this page. These will be discussed in detail later in the tutorial.

Welcome to Manatee

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

ACCESS LISTINGS

Annotation Tools

Genome Summary

Gene Ontology

Genome Properties

Genome Viewer

Multi Genome Annotation Tool

submit

reset

ACCESS GENE CURATION PAGE

gene:

SEARCH GENES BY GENE NAME

gene name:

CHANGE ORGANISM DATABASE

database:

Refresh XML

Search

Asmbl Id: 7974

Database: gsp

database:

asmbl\_id:

submit

reset

| feat_name | end5 | end3  | role_id | ec_num   | gene_sym | complete | com_name                             |
|-----------|------|-------|---------|----------|----------|----------|--------------------------------------|
| ORF02394  | 954  | 2504  | 142     |          |          | complete | proton/peptide symporter family prot |
| prop_acc  |      | state |         | property |          |          |                                      |
|           |      |       |         |          |          |          |                                      |

Six Frame Options (on six frame clicks)

View Sequence

Blast

Insert Gene

Gene Options (on feat\_name clicks)

View Sequence

Annotate ORF

Edit Start

Extend Gene

Blast ORF

Merge Genes

Delete Gene

0.0kb1.0kb2.0kb3.0kb4.0kb5.0kb6.0kb7.0kb

ORF02394

ORF02395

ORF02393

ORF02392

ORF02391

ORF08000

-->

Search

Coordinate:

Search

Lower Coordinate:

Upper Coordinate:

Search

feat\_name:

Search

26

# The Welcome to Manatee page, links from Access Listings:

## “Multi Genome Annotation Tool” (MGAT)

The MGAT tool allows the annotation of orthologous genes from several genomes at one time. It is linked into Manatee at several points. MGAT is still undergoing development and is not currently available for public use. A separate tutorial for this tool is under construction.

### Welcome to Manatee

Welcome to Manatee

2 This is the main menu page for the Manatee tool. One can access genes directly by their gene id number or name or link to

ACCESS LISTINGS

- Annotation Tools
- Genome Summary
- Gene Ontology
- Genome Properties
- Genome Viewer
- Multi Genome Annotation Tool

submit

reset

ACCESS GENE CURATION PAGE

SEARCH GENES BY GENE NAME

CHANGE ORGANISM DATABASE

database:

Welcome to MGAT

Access COGS By Role Categories

Single Role Category:

Main Role Category:

Submit

Unclassified  
Amino acid biosynthesis  
Purines, pyrimidines, nucleosides, and nucleotides  
Fatty acid and phospholipid metabolism  
Biosynthesis of cofactors, prosthetic groups, and carriers  
Central intermediary metabolism  
Energy metabolism  
Transport and binding proteins  
DNA metabolism  
Transcription  
Protein synthesis  
Protein fate

| Ident Information           |                          |                                                                                                                                                                       |        |          |         |                                                            |
|-----------------------------|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|----------|---------|------------------------------------------------------------|
| AllV All align              | manatee                  | com_name                                                                                                                                                              | ec_num | gene_sym | role_id | mainrole                                                   |
| <a href="#">BMA3090</a>     | <a href="#">ORF04141</a> | thiamine biosynthesis protein ThiC                                                                                                                                    |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">GHNORF3126</a>  | <a href="#">ORF03083</a> | thiamin biosynthesis protein thiC                                                                                                                                     |        | THIC     | 126     | Purines, pyrimidines, nucleosides, and nucleotides         |
| <a href="#">AFE2099</a>     | <a href="#">ORF03389</a> | thiamine biosynthesis protein ThiC                                                                                                                                    |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">CC2029</a>      | <a href="#">ORF06212</a> | thiamine biosynthesis protein ThiC                                                                                                                                    |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">NT02NM2427</a>  | <a href="#">ORF03627</a> | thiamine biosynthesis protein thiC. {neisseria meningitidis serogroup a;} ^1^pirk81956k81956 thiamin biosynthesis protein nma0397 [imported] - neisseria meningitidis |        |          | 185     | Unclassified                                               |
| <a href="#">NMB2040</a>     | <a href="#">ORF01381</a> | thiamine biosynthesis protein ThiC                                                                                                                                    |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">MXAN4235</a>    | <a href="#">ORF01813</a> | thiamine biosynthesis protein ThiC                                                                                                                                    |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">PSPPH0535</a>   | <a href="#">ORF00554</a> | thiamin biosynthesis protein ThiC                                                                                                                                     |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">NT01PS1691</a>  | <a href="#">ORF02811</a> | thiamin biosynthesis protein ThiC                                                                                                                                     |        | thiC     | 126     | Purines, pyrimidines, nucleosides, and nucleotides         |
| <a href="#">PSPT04976</a>   | <a href="#">ORF04688</a> | thiamin biosynthesis protein ThiC                                                                                                                                     |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">PF5841</a>      | <a href="#">ORF08667</a> | thiamin biosynthesis protein ThiC                                                                                                                                     |        |          | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">PP4922</a>      | <a href="#">ORF01030</a> | thiamin biosynthesis protein ThiC                                                                                                                                     |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |
| <a href="#">GEBCORF0191</a> | <a href="#">ORF00186</a> | Thiamine biosynthesis protein thiC                                                                                                                                    |        | thiC     | 162     | Biosynthesis of cofactors, prosthetic groups, and carriers |

### MGAT Search information

database :

**Biosynthesis of cofactors, prosthetic gro**

| Genome Annotation        |                                                                              |         |      |       |                   |                           |                         |          |
|--------------------------|------------------------------------------------------------------------------|---------|------|-------|-------------------|---------------------------|-------------------------|----------|
| gene                     | com_name                                                                     | SP Fams | Cogs | JFams | matrix            | TigrFam                   | PFAM                    | start_ec |
| <a href="#">ORF05157</a> | thiamin biosynthesis protein ThiC                                            |         |      |       | <a href="#">X</a> | <a href="#">TIGR00190</a> |                         | No       |
| <a href="#">ORF05164</a> | thiH protein                                                                 |         |      |       | <a href="#">X</a> | <a href="#">TIGR02351</a> |                         | No       |
| <a href="#">ORF03392</a> | thiH protein, putative                                                       |         |      |       | <a href="#">X</a> |                           |                         | No       |
| <a href="#">ORF01112</a> | ApbE family protein, putative                                                |         |      |       |                   |                           | <a href="#">PF02424</a> | No       |
| <a href="#">ORF05159</a> | phosphomethylpyrimidine kinase/thiamin-phosphate pyrophosphorylase, putative |         |      |       |                   | <a href="#">TIGR00693</a> |                         | No       |

# The Welcome to Manatee Page: Options under “Annotation Tools”

Welcome to Manatee

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 a

ACCESS LISTINGS

- Annotation Tools
- Genome Summary
- Gene Ontology
- Genome Properties
- Genome Viewer
- Multi Genome Annotation Tool

submit

reset

ACCESS GENE CURATION PAGE

gene:

SEARCH GENES BY GENE NAME

gene name:

CHANGE ORGANISM DATABASE

database:

Descriptions of a few of the links and tools on this page are described here in the red boxes, following slides will detail the other options/tools on this page.

Shewanella oneidensis MR-1 Annotation Tools Home | Logged into [gsp] as mlgwin

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.

Home Annotation Tools Genome Summary Gene Assignment Help

ACCESS GENE LISTS

molecule: All molecules

all genes, ordered by role category (extended hypothetical protein)

main role category Unclassified

single role category role\_id

ACCESS GENE CURATION PAGE

gene: This is the same option as the one on the “Genome Summary” page

ACCESS GENES BY COORDINATE RANGE

end5: end3:

OVERLAP ANALYSIS

location: /usr/local/annotation/GSP/overlap\_analysis/gsp.overlap

INTEREVIDENCE REGION ANALYSIS

location: /usr/local/annotation/GSP/intergenic\_regions/gsp.REPC

CUSTOM QUERY

type a custom sql query here to see a list of genes with criteria not available on gene list options

OTHER TOOLS

- Data Consistency Checks
- Frameshift Reports
- Hypothetical Protein List
- Annotation Status
- Phage Region Viewer
- PubMed Organism Search

# "Annotation Tools": "Coordinate Range"

► List of all genes found between 10000 - 20000

| A | C | gene id   | locus   | end5  | end3  | role id  | gene name                                     | gene symbol | ec       |
|---|---|-----------|---------|-------|-------|----------|-----------------------------------------------|-------------|----------|
|   |   | ORF02375  | SO0017  | 22090 | 18941 | 156      | conserved hypothetical protein                |             |          |
|   |   | ORF02378  | SO0016  | 18279 | 18854 | 132      | DNA-3-methyladenine glycosidase I             | tag         | 3.2.2.20 |
|   |   | ORF02379  | SO0015  | 18161 | 17256 | 137      | glycyl-tRNA synthetase, alpha subunit         | glyQ        | 6.1.1.14 |
|   |   | ORF02381  | SO0014  | 17246 | 15180 | 137      | glycyl-tRNA synthetase, beta subunit          | glyS        | 6.1.1.14 |
|   |   | ORF02382  | SO0013  | 14311 | 15111 |          | hypothetical protein                          |             |          |
|   |   | ORF02383  | SO0012  | 13791 | 13129 | 96 102,  | glutathione S-transferase family protein      |             |          |
|   |   | ORF02385  | SO0011  | 10638 | 13055 | 132      | DNA gyrase, B subunit                         | gyrB        | 5.99.1.3 |
|   |   | ORF02386  | SO0010  | 9539  | 10621 | 132      | DNA replication and repair protein RecF       | recF        |          |
|   |   | ORFA00005 | SOA0024 | 20332 | 19523 | 154      | ISSo1, transposase OrfB                       |             |          |
|   |   | ORFA00006 | SOA0023 | 19154 | 19453 | 94 186,  | proteic killer suppressor protein             | higA        |          |
|   |   | ORFA00007 | SOA0022 | 18774 | 19079 | 94 186,  | proteic killer active protein                 | higB        |          |
|   |   | ORFA00008 | SOA0021 | 18235 | 18462 | 154 270, | ISSo1, transposase OrfB, truncation           |             |          |
|   |   | ORFA00009 | SOA0020 | 17414 | 18154 | 154 270, | transposase family protein, truncation        |             |          |
|   |   | ORFA00011 | SOA0019 | 16733 | 17290 | 132 154, | TnSon1, resolvase                             |             |          |
|   |   | ORFA00012 | SOA0018 | 16362 | 16739 | 154 156, | TnSon1, conserved hypothetical protein        |             |          |
|   |   | ORFA00013 | SOA0017 | 16075 | 16365 | 703      | TnSon1, nucleotidyltransferase domain protein |             |          |
|   |   | ORFA00014 | SOA0016 | 15911 | 12945 | 154      | TnSon1, transposase                           |             |          |
|   |   | ORFA00015 | SOA0015 | 12878 | 12732 |          | hypothetical protein                          |             |          |
|   |   | ORFA00016 | SOA0014 | 12332 | 12427 |          | hypothetical protein                          |             |          |
|   |   | ORFA00017 | SOA0013 | 11739 | 11335 | 132      | umuD protein                                  | umuD        | 3.4.2.2  |
|   |   | ORFA00019 | SOA0012 | 11334 | 10078 | 132      | umuC protein                                  | umuC        |          |

Input a coordinate range and you will get a list of genes whose coordinates fall anywhere in that range.

# “Annotation Tools”: “Overlap Analysis”

**Shewanella oneidensis MR-1** | **Annotation Tools** | [Home](#) | Logged into [gsp] as mlgwin

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

Home | **Annotation Tools** | Genome Summary | Gene Assignment Help

**ACCESS GENE LISTS**

molecule: All molecules

☒ all genes, ordered by role category (excludes hypothetical proteins)

☐ main role category: Unclassified

☐ single role category: role\_id

**ACCESS GENE CURATION PAGE**

gene:

**ACCESS GENES BY COORDINATE RANGE**

end5: end3:

**OVERLAP ANALYSIS**

location: /usr/local/annotation/GSP/overlap\_analysis/gsp.overlap

**INTEREVIDENCE REGION ANALYSIS**

location: /usr/local/annotation/GSP/intergenic\_regions/gsp.REPC

**CUSTOM QUERY**

submit reset

**OTHER TOOLS**

- Data Consistency Checks
- Frameshift Reports
- Hypothetical Protein List
- Annotation Status
- Phage Region Viewer
- PubMed Organism Search

We work on the premise that genes do not generally overlap in prokaryotic genomes. We look for overlapping genes predicted by Glimmer and where we can, remove genes suspected of being false calls by Glimmer. Often overlap between two genes can be resolved by the curation of the start site of one or both, or by the removal of a “hypothetical protein” (one that has no similarity to anything) when it overlaps a protein with very clear similarity to other proteins. For more on overlap analysis see the Annotation Overview document. This display shows the pairs of overlapping genes as indicated by the background color shifts from blue to white to blue to white. Clicking on the feature id number takes you to a Gene Curation Page (GCP) for that gene. Also displayed are the percent of overlap, name of the protein, and notes from Glimmer regarding the protein in question.

**Shewanella oneidensis MR-1** | **Overlap Display** | [Home](#) | Logged into [gsp] as mlgwin

This page displays a list of overlapping ORFs that have been identified as candidate genes by GLIMMER (the gene finding program).

| feature   | % overlap | name                                                     | glimmer                                   |
|-----------|-----------|----------------------------------------------------------|-------------------------------------------|
| ORFA00023 | 3.47      | hypothetical protein                                     | 742 [-3 L= 759] [DelayedByORFA00024 L=45] |
| ORFA00022 | 8.77      | hypothetical protein                                     | 229 [-2 L= 282]                           |
| ORFA00078 | 6.41      | ISSo8, transposase, degenerate                           | 423 [-1 L= 543]                           |
| ORFA00082 | 4.38      | transposase, IS3 family, degenerate                      | 254 [+3 L= 336]                           |
| ORFA00096 | 2.60      | ISSo12, transposase                                      | 672 [+1 L= 921]                           |
| ORFA00095 | 19.05     | hypothetical protein                                     | 97 [-2 L= 123]                            |
| ORFA00204 | 3.92      | conserved hypothetical protein                           | 99 [-3 L= 108]                            |
| ORFA00203 | 4.23      | hypothetical protein                                     | 812 [-1 L= 1128]                          |
| ORF00040  | 6.46      | TonB-dependent receptor C-terminal region domain protein | 198 [-1 L= 477] [DelayedByORFA00041 L=18] |
| ORF00041  | 14.55     | hypothetical protein                                     | 197 [-2 L= 210]                           |
| ORF00158  | 5.00      | hypothetical protein                                     | 215 [-3 L= 537]                           |
| ORF00157  | 3.33      | ISSo1, transposase OrfB                                  | 693 [+1 L= 807]                           |
| ORF00177  | 5.23      | hypothetical protein                                     | 297 [+2 L= 360]                           |
| ORF00176  | 3.15      | conserved hypothetical protein                           | 396 [+1 L= 645]                           |

# “Annotation Tools”: “Interevidence Analysis”

Glimmer is known to sometimes miss identifying a few real genes. This is especially true for areas of the genome that have been laterally transferred. To find genes Glimmer might have missed, we run an analysis called “interevidence”. This tool takes the nucleotide sequence between genes, the sequence of hypothetical proteins (those that have similarity to nothing), and any regions of proteins that have similarity to nothing, does a 6 frame translation, and then searches those translations against niaa (our in-house protein db). Any possible areas of similarity are then reviewed by annotators and missed genes are entered into the db.

Shewanella oneidensis MR-1

Annotation Tools

Home | Logged into [gsp] as mlgwin

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

HomeAnnotation ToolsGenome SummaryGene Assignment Help

ACCESS GENE LISTS

molecule: All molecules

all genes, ordered by role category (excludes hypothetical proteins)

main role category: Unclassified

single role category: role\_id

ACCESS GENE CURATION PAGE

gene:

ACCESS GENES BY COORDINATE RANGE

end5: end3:

OVERLAP ANALYSIS

location: /usr/local/annotation/GSP/overlap\_analysis/gsp.overlap

INTEREVIDENCE REGION ANALYSIS

location: /usr/local/annotation/GSP/intergenic\_regions/gsp.REPC

CUSTOM QUERY

OTHER TOOLS

Data Consistency Checks

Frameshift Reports

Hypothetical Protein List

Annotation Status

Phage Region Viewer

PubMed Organism Search

Shewanella oneidensis MR-1

Interevidence Regions Display

Home | Logged into [gsp] as mlgwin

This page displays the output from the Interevidence Region analysis program. Areas of the genome with no genes are searched for homology in case genes were missed by GLIMMER.

| IE gene             | asmb_id | IE end5 | IE end3                                           | Hit end5 | Hit end3 | Proposed ORF end5 | Proposed ORF end3 |
|---------------------|---------|---------|---------------------------------------------------|----------|----------|-------------------|-------------------|
| gsp_7971_R2833_orf1 | 3       | 383     | :                                                 |          |          |                   |                   |
| gsp_7971_R2833_orf2 | 3       | 167     | :                                                 |          |          |                   |                   |
| gsp_7971_R2833_orf3 | 3       | 383     | :                                                 |          |          |                   |                   |
| gsp_7971_R2832_orf2 | 293     | 1084    | :                                                 |          |          |                   |                   |
| gsp_7971_R2832_orf1 | 347     | 1084    | :                                                 |          |          |                   |                   |
| gsp_7971_R3_orf2    | 1271    | 1441    | :                                                 |          |          |                   |                   |
| gsp_7971_R3_orf1    | 1345    | 2010    | ORF00003: hypothetical protein                    |          |          |                   |                   |
| gsp_7971_R2_orf1    | 1424    | 924     | :                                                 |          |          |                   |                   |
| gsp_7970_R8_orf1    | 4285    | 3449    | ORFA00009: transposase family protein, truncation |          |          |                   |                   |
| gsp_7970_R23_orf1   | 30648   | 30881   | :                                                 |          |          |                   |                   |

58.2/72.8% over 54aa

Mesorhizobium loti

GPI4026131dbj transposase

Insert characterized

gsp\_7971\_R2833\_orf1(1.3 - 54.3 of 127 aa)

GP|I4026131|dbj|BAB52729.1|AP003009(288 - 341 of 397) transposase (Mesorhizobium loti)

%Match = 8.5

%Identity = 58.2 %Similarity = 72.7

Matches = 32 Mismatches = 13 Conservative Sub.s = 8

Gaps = 2 InDels = 6 Frame Shifts = 0

Primary Frame = 2 [0, 53, 0]

agtagcgggtgtagtccagggtcgctttgttggtggccttagaggccagctctgactaccagttttcgccac atcatg

catcactccttcgtccgggtttgtatcggtgccgtatcagtgaggctgaaactaccacagctgagagaaaaa tgtgga

cctttttgaagacaaactagttccaaattactaacctgactcaccgtcgctcagcggcagtgacactctgtgtga ggactt

-55.7 -45.7 -35.7 -25.7 -15.7 -5.7 5.3 14.3

TDFTHPVASVLASVSRRMVVCLDLSCGLSAAVDLS\*SVSEGGPLSEYH\*AMQSTQPKACLCYRDGQHQH\*MWLSCD

DRRAFLRLHAPVRRKRWVVYAKAPFAGPEAVLAYLSRYTHRVAISNSRLIRLDESGVTFRYKDYRRDGADRQQLMTLGPD

240 250 260 270 280 290 300

gtactcttcgtagcagcagctgctgagtcacacgcacccccaccggtagctgaagttttctctgctctagtaacctgg

attggattatcagttgtgatttcacgggaattctttgatgacatttcgctaaagtgccgagatgatattgtgacctt

ctccttagtgagcaggagattttcttcacgggcccgaatggggaggttggtgctggatgacgattgttggaagggag

24.3 34.3 44.3 54.3 64.3 74.3 84.3 94.3

EFIRRYLLHVLPKGLMRVRHFGFLANACRRKKLALILRLRKPQVVLASPLVKKDCLWSCPQCQLGHLQFIGLIRPQSVV

EFIRRFLLHVLPGFERRHRCGLLAGSARKASLALARELLNVAVPSPDDDTTNEPDDFRPPCCCGRMIVIEVFERWRQ

320 330 340 350 360 370 380

31

# “Annotation Tools”: “Other Tools” section

Shewanella oneidensis MR-1

Annotation Tools

Home | Logged into [gsp] as mlgwinn

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 as a resource for locating general properties of the genome and determining the progress made in the Annotation of the genome of interest.

Home

Annotation Tools

Genome Summary

Gene Assignment Help

ACCESS GENE LISTS

molecule:

All molecules

all genes, ordered by role category (excludes hypothetical proteins)

main role category

Unclassified

single role category

role\_id

ACCESS GENE CURATION PAGE

gene:

ACCESS GENES BY COORDINATE RANGE

end5:

end3:

OVERLAP ANALYSIS

location:

/usr/local/annotation/GSP/overlap\_analysis/gsp.overlap

INTEREVIDENCE REGION ANALYSIS

location:

/usr/local/annotation/GSP/intergenic\_regions/gsp.REPC

CUSTOM QUERY

submit

reset

OTHER TOOLS

Data Consistency Checks

Frameshift Reports

Hypothetical Protein List

Annotation Status

Phage Region Viewer

PubMed Organism Search

Data consistency checks:

Clicking this generates a list of possible errors or consistency problems in the annotation. For example, if two proteins have the same common name but different TIGR role assignments, they would be listed in the consistency check section for review.

Frameshift Reports:

Similar to the “Frameshift status” link that was described earlier for the “Genome Summary” page - basically a list of genes with frameshift reports to be resolved.

Hypothetical protein list:

a list of hypothetical proteins, (those with insufficient evidence to make any functional assignment) for which there is any shred of information which might lead to annotation other than “hypothetical protein”, this list is generated automatically after AutoAnnotate has made its initial assignment. Those “hypothetical proteins” called by AutoAnnotate that have any BER or HMM evidence are put on this list for manual review.

Annotation status:

The same page as was described from the “Genome Summary” page - lists of the steps in annotation and a list of role categories, status of completion and annotator who did the work is noted.

Phage Region Viewer:

A tool that lists any identified prophage regions in the genome and the genes within them.

PubMed Organism Search:

Automatically takes you to the NCBI PubMed site and gives results for a PubMed search using the organism name as keywords. Useful for finding 32 literature on the organism you are working on.

# “Annotation Tools”: “Access Gene Lists” section

Although all of the tools described so far in this tutorial are quite useful, the bulk of annotator time is spent in viewing and editing information that is displayed on gene lists and Gene Curation Pages that are accessed through the “Access Gene Lists” section.

This tool will create a table of genes chosen according to the options in the red box at right. As mentioned in the overview, at TIGR we organize our annotation efforts around TIGR role categories. This tool allows us to view the genes within each TIGR role category.

The first option to select in this section is which molecule you wish to annotate. Some genomes consist of just one chromosome and nothing else, while others can have multiple chromosomes or chromosome(s) and one or more plasmids. If multiple DNA molecules exist for the genome in question, the pull down menu at the top of this section will list them along with their id number. The default selection is “All molecules” as the team usually annotates all molecules at once, however, to choose just one of the molecules, simply select it from the pull-down menu.

Then choose one of the 3 options for which role categories you want to see genes from with the toggle buttons: first you can choose all role categories, second you can choose one particular main role category, and third you can choose one particular sub-role category. All of the mainrole categories are listed in the pull-down menu in the main role category selection, to choose one, simply highlight it. In order to select a particular sub-role category you must enter into the box next to “single role category” the id number of the sub-role category. There is a listing of all of the TIGR role categories and their id numbers on the next two pages of this tutorial.

Once you have chosen your desired options, click submit to see a list of the genes that fit your selections.

Shewanella oneidensis MR-1

Annotation Tools

Home | Logged into [gsp] as mlgwinn

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 as a resource for locating general properties of the genome and determining the progress made in the Annotation of the genome of interest.

Home

Annotation Tools

Genome Summary

Gene Assignment Help

ACCESS GENE LISTS

▶ molecule: All molecules

all genes, ordered by role category (excludes hypothetical proteins)

main role categoryUnclassified

single role categoryrole\_id

ACCESS GENE CURATION PAGE

▶ gene:

ACCESS GENES BY COORDINATE RANGE

▶ end5: end3:

OVERLAP ANALYSIS

▶ location: /usr/local/annotation/GSP/overlap\_analysis/gsp.overlap

INTEREVIDENCE REGION ANALYSIS

▶ location: /usr/local/annotation/GSP/intergenic\_regions/gsp.REPC

CUSTOM QUERY

OTHER TOOLS

▶ Data Consistency Checks

▶ Frameshift Reports

▶ Hypothetical Protein List

▶ Annotation Status

▶ Phage Region Viewer

▶ PubMed Organism Search

submit

reset

33

**Gene List:** The results of your selection from the Access Listings tool are displayed in a gene list containing gene id number, locus (if available), coordinates of the gene (end5, end3), common name of the gene/protein, gene\_sym, EC number, and other roles for the protein. Not all of these fields will be populated for every gene. The genes are organized by role category (if your selection included more than one.) There are many features of the gene list, and much information displayed - text describing a feature is boxed in the same color as the feature itself.

Clicking on the blue names of any mainrole category takes you to a gene list for that category.

*Shewanella oneidensis* MR-1    Gene List

This List contains ORFs which are currently assigned to TIGR microbial role categories. It is sorted by role category.

- ▶All categories   ▶Unclassified   ▶Amino acid biosynthesis   ▶Purines, pyrimidines, nucleosides, and nucleotides   ▶Fatty acid and phospholipid metabolism   ▶Biosynthesis of cofactors, prosthetic groups, and carriers   ▶Central intermediary metabolism   ▶Energy metabolism   ▶Transport and binding proteins   ▶DNA metabolism   ▶Transcription   ▶Protein synthesis   ▶Protein fate   ▶Regulatory functions   ▶Signal transduction   ▶Cell envelope   ▶Cellular processes   ▶Mobile and extrachromosomal element functions   ▶Unknown function   ▶Hypothetical proteins   ▶Disrupted reading frame   ▶Glimmer rejects

**Biosynthesis of cofactors, prosthetic groups, and carriers**

View list of Genome Properties found for this role category

| ▶Biotin     |   | Role id: 77   |        |                     |         | Edit Annotation Notebook                                    |             |          |             |
|-------------|---|---------------|--------|---------------------|---------|-------------------------------------------------------------|-------------|----------|-------------|
| prop_acc    |   | state         |        | property            |         |                                                             | is public?  |          |             |
| GenProp0036 |   | YES           |        | biotin biosynthesis |         |                                                             | YES         |          |             |
| A           | C | gene id       | locus  | end5                | end3    | gene name                                                   | gene symbol | ec       | other roles |
|             |   | ORF02146 (GV) | SO0214 | 224657              | 225616  | birA bifunctional protein                                   | birA        | 6.3.4.15 | 129         |
|             |   | ORF02395 (GV) | SO0001 | 774                 | 334     | mioC protein                                                | mioC        |          | 112         |
|             |   | ORF02552 (GV) | SO4626 | 4821460             | 4822251 | bioH protein                                                | bioH        |          |             |
|             |   | ORF04812 (GV) | SO2741 | 2856886             | 2858271 | adenosylmethionine--8-amino-7-oxononanoate aminotransferase | bioA        | 2.6.1.62 |             |
|             |   | ORF04813 (GV) | SO2740 | 2856763             | 2855711 | biotin synthase                                             | bioB        | 2.8.1.6  |             |
|             |   | ORF04814 (GV) | SO2739 | 2855489             | 2854284 | 8-amino-7-oxononanoate synthase                             | bioF        | 2.3.1.47 |             |
|             |   | ORF04816 (GV) | SO2738 | 2854163             | 2853336 | biotin synthesis protein BioC                               | bioC        |          |             |
|             |   | ORF04817 (GV) | SO2737 | 2853281             | 2852586 | dethiobiotin synthase                                       | bioD        | 6.3.3.3  |             |

A green dot in the "A" column indicates this orf was given a high quality assignment by AutoAnnotate. (The only type of evidence that will currently trigger this is an above trusted cutoff hit to an equivalog HMM.) A pink dot will appear in the "C" column once an annotator has finished annotation for the gene and marked it complete

Link to role notes for this category

Click on the gene\_id (feat\_name) link to see the Gene Curation Page for each gene. Click on "GV" for Genome Viewer.

This links to a text entry field to store info of interest to the project that is found during annotation.

The ORFs can be ordered according to any of the blue headers by clicking on that header.

# TIGR Role Categories - Page 1

Unclassified (the automated program was unable to assign a role to these)

185 Role category not yet assigned

## Amino acid biosynthesis

- 70 Aromatic amino acid family
- 71 Aspartate family
- 73 Glutamate family
- 74 Pyruvate family
- 75 Serine family
- 161 Histidine family
- 69 Other

## Purines, pyrimidines, nucleosides, and nucleotides

- 123 2'-Deoxyribonucleotide metabolism
- 124 Nucleotide and nucleoside interconversions
- 125 Purine ribonucleotide biosynthesis
- 126 Pyrimidine ribonucleotide biosynthesis
- 127 Salvage of nucleosides and nucleotides
- 128 Sugar-nucleotide biosynthesis and conversions
- 122 Other

## Fatty acid and phospholipid metabolism

- 176 Biosynthesis
- 177 Degradation
- 121 Other

## Biosynthesis of cofactors, prosthetic groups, and carriers

- 77 Biotin
- 78 Folic acid
- 79 Heme, porphyrin, and cobalamin
- 80 Lipoate
- 81 Menaquinone and ubiquinone
- 82 Molybdopterin
- 83 Pantothenate and coenzyme A
- 84 Pyridoxine
- 85 Riboflavin, FMN, and FAD
- 86 Glutathione
- 162 Thiamine
- 163 Pyridine nucleotides
- 191 Chlorophyll
- 707 Siderophores
- 76 Other

## Central intermediary metabolism

- 100 Amino sugars
- 698 One-carbon metabolism
- 103 Phosphorus compounds
- 104 Polyamine biosynthesis
- 106 Sulfur metabolism
- 179 Nitrogen fixation
- 160 Nitrogen metabolism
- 709 Electron carrier regeneration
- 102 Other

## Energy metabolism

- 108 Aerobic
- 109 Amino acids and amines
- 110 Anaerobic
- 111 ATP-proton motive force interconversion
- 112 Electron transport
- 113 Entner-Doudoroff
- 114 Fermentation
- 116 Glycolysis/gluconeogenesis
- 117 Pentose phosphate pathway
- 118 Pyruvate dehydrogenase
- 119 Sugars
- 120 TCA cycle
- 159 Methanogenesis
- 105 Biosynthesis and degradation of polysaccharides
- 164 Photosynthesis
- 180 Chemoautotrophy
- 184 Other

## Transport and binding proteins

- 142 Amino acids, peptides and amines
- 143 Anions
- 144 Carbohydrates, organic alcohols, and acids
- 145 Cations and iron carrying compounds
- 146 Nucleosides, purines and pyrimidines
- 182 Porins
- 147 Other
- 141 Unknown substrate

# TIGR Role Categories - Page 2

## DNA metabolism

|     |                                            |
|-----|--------------------------------------------|
| 132 | DNA replication, recombination, and repair |
| 183 | Restriction/modification                   |
| 131 | Degradation of DNA                         |
| 170 | Chromosome-associated proteins             |
| 130 | Other                                      |

## Transcription

|     |                              |
|-----|------------------------------|
| 134 | Degradation of RNA           |
| 135 | DNA-dependent RNA polymerase |
| 165 | Transcription factors        |
| 166 | RNA processing               |
| 133 | Other                        |

## Protein synthesis

|     |                                                |
|-----|------------------------------------------------|
| 137 | tRNA aminoacylation                            |
| 158 | Ribosomal proteins: synthesis and modification |
| 168 | tRNA and rRNA base modification                |
| 169 | Translation factors                            |
| 136 | Other                                          |

## Protein fate

|     |                                                      |
|-----|------------------------------------------------------|
| 97  | Protein and peptide secretion and trafficking        |
| 140 | Protein modification and repair                      |
| 95  | Protein folding and stabilization                    |
| 138 | Degradation of proteins, peptides, and glycopeptides |
| 189 | Other                                                |

## Regulatory functions

|     |                             |
|-----|-----------------------------|
| 261 | DNA interactions            |
| 262 | RNA interactions            |
| 263 | Protein interactions        |
| 264 | Small molecule interactions |
| 129 | Other                       |

## Signal transduction

|     |                       |
|-----|-----------------------|
| 699 | Two-component systems |
| 700 | PTS                   |
| 710 | Other                 |

## Cell envelope

|    |                                                                                 |
|----|---------------------------------------------------------------------------------|
| 91 | Surface structures                                                              |
| 89 | Biosynthesis of murein sacculus and peptidoglycan                               |
| 90 | Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides |
| 88 | Other                                                                           |

## Cellular processes

|     |                                    |
|-----|------------------------------------|
| 93  | Cell division                      |
| 188 | Chemotaxis and motility            |
| 701 | Cell adhesion                      |
| 702 | Conjugation                        |
| 96  | Detoxification                     |
| 98  | DNA Transformation                 |
| 705 | Sporulation and Germination        |
| 94  | Toxin production and resistance    |
| 187 | Pathogenesis                       |
| 149 | Adaptations to atypical conditions |
| 706 | Bioosynthesis of natural products  |
| 92  | Other                              |

## Mobile and extrachromosomal element functions

|     |                      |
|-----|----------------------|
| 186 | Plasmid functions    |
| 152 | Prophage functions   |
| 154 | Transposon functions |
| 708 | Other                |

## Unknown

|     |                                |
|-----|--------------------------------|
| 703 | Enzymes of unknown specificity |
| 157 | General                        |

## Hypothetical

|     |           |
|-----|-----------|
| 156 | Conserved |
| 704 | Domain    |

## Disrupted reading frame

|     |      |
|-----|------|
| 270 | NULL |
|-----|------|

## Gene list link: Edit Annotation Notebook:

Clicking on the “Edit Annotation Notebook” link on the gene list page will take you to a page where you can enter or edit annotation notes for a particular role category. It is in this text field that we store information that we think will be useful for the PI of the project in the analysis of the genome or in the preparation of the manuscript. Things such as the presence of an unexpected pathway, or the fact that a key step in another pathway is missing. Once the text is as you want it, click “submit” to store the information in the db.

|                                                                                                                                                                                                                                                                                                         |  |                                      |                                             |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--------------------------------------|---------------------------------------------|
| <b><i>Shewanella oneidensis</i> MR-1</b>                                                                                                                                                                                                                                                                |  | <b>Annotation Notes - role id 83</b> | Logged into [gsp] as <a href="#">mlgwin</a> |
|  The annotation_notebook.txt script directs the user to a web display page that contains annotators' comments about particular genes or regions of the genome that the annotators thought were unusual or interesting. |  |                                      |                                             |
| <div>Appears to lack panD which I searched for with the E.coli sequence. No matches using blastp or tblastn. RTD</div>                                                                                                                                                                                  |  |                                      |                                             |
| <input type="button" value="Update"/>                                                                                                                                                                                                                                                                   |  | <input type="button" value="Reset"/> |                                             |

## Gene list link: Role information page:

TIGR annotators expert in particular role categories have written “role notes” to aid new annotators and annotators unfamiliar with the category in the annotation process. These notes contain information on what genes belong in the category and what genes don't, on the pathways found in particular categories, and on the TIGR naming conventions for proteins within the category.

Any TIGR annotator can update or add text to the note field by typing it in and then clicking submit.

There is also a link to the role notes pages from the Gene Curation Page (GCP) which will be shown in the GCP section.

| Shewanella oneidensis MR-1                                                                                                                                                                                                             | Role Information For Role_id 77 |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|
|  The role_info.cgi script is executed from the Submit web display page and directs the user to a web display page that contains Single Role Category. |                                 |

### Role 77 Biosynthesis of cofactors, prosthetic groups, and carriers - Biotin

#### Role Info:

```
Genes involved in the synthesis of biotin.

pathway from 6-carboxyhexanoyl-CoA plus L-alanine to biotin:
step  gene
1      8-amino-7-oxononanoate synthase (bioF)
TIGR00858: bioF
2      adenosylmethionine-8-amino-7-oxononanoate aminotransferase
(bioA)
TIGR00508: bioA
3      dethiobiotin synthetase (bioD)
TIGR00347: bioD
4      biotin synthase (bioB)
TIGR00433: bioB
Other genes also involved:

BirA bifunctional protein (birA)
    acts as operon repressor, synthesizes corepressor, activates
    biotin,
    and transfers activated biotin to proteins
biotin synthesis protein BioC (bioC)
    involved in an early, undefined step in biotin synthesis
biotin sulfoxide reductase (bisZ)
    changes biotin sulfoxide back to biotin, scavenging reaction
TIGR01738 bioH protein (bioH)
    in early steps of biotin biosynthesis
TIGR01204 bioW protein = 6-carboxyhexanoate--CoA ligase
    found in Bacillus and Methanococcus, involved in biotin
    synthesis
    BioW plus BioF of Bacillus can replace bioC and bioH of E.
    coli
    (says PMID:2110099)

In many, but by no means all, organisms most of these genes can be
found in an operon.

mioC protein: MioC is a flavodoxin thought to function as an electron
transporter (role_id=112) and in biotin biosynthesis (role_id=77).
mioC neighbors oriC in E. coli. Early studies on mioC expression
demonstrate a dramatic effect on initiation of chromosome duplication
at oriC on minichromosomes. This role has not been demonstrated in
duplication of the wild type chromosome. Additionally, the
minichromosome is not necessarily a valid model for chromosomal
replication. Because of this dubious association with chromosomal
```

[submit](#) Update Role Note For 77

# Gene Curation Page

The Gene Curation Page (GCP) is likely the most important page within Manatee, it is certainly the one that annotators spend the bulk of their time looking at and working with.

This page can be accessed within Manatee from many places:  
any gene list, the “Access Gene Curation Page” option on the Genome Summary/Annotation Tools pages, Genome Viewer, .... and more.

The GCP is a very complex page so we will look at it in sections. I will try to organize the descriptions of each section in roughly the same order that the concepts behind each section were reviewed in the Annotation Overview.

| Shewanella oneidensis MR-1                                                                                                      |  | Gene Curation Page                                                                                                               |  | <a href="#">Home</a>   <a href="#">Logged into [gsp] as mlgwin</a>                                                               |  |
|---------------------------------------------------------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------------------------------------------|--|
|                                                |  |                                                                                                                                  |  |                                                                                                                                  |  |
| <b>GENE CURATION INFORMATION</b>             |  |                                                                                                                                  |  |                                                                                                                                  |  |
| <b>ORF04813 (SO2740)</b><br><a href="#">View BER Searches</a><br>asmb_id: 7974<br><br><a href="#">Reload Page</a>               |  | end5/end3: 2856763 / 2855711<br>gene length: 1053<br>protein length: 350<br>molecular wt: 38790.13                               |  | database: <input type="text" value="gsp"/><br>feat_name / locus: <input type="text"/><br><input type="button" value="New Gene"/> |  |
| <input type="text" value="Select Display"/>  |  | <input type="text" value="Select Function"/>  |  | <input type="button" value="Refresh Searches"/>                                                                                  |  |

| <b>GENE IDENTIFICATION</b>                                                                                                                |  | <a href="#">submit</a>   <a href="#">history</a>    |  |
|-------------------------------------------------------------------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------------------------------------------------|--|
| gene name:<br><input type="text" value="biotin synthase"/>                                                                                |  |                                                                                                                                        |  |
| gene_sym:<br><input type="text" value="bioB"/>                                                                                            |  |                                                                                                                                        |  |
| EC number(s):<br><input type="text" value="2.8.1.6"/>  |  | EC GO suggestions:<br><a href="#">GO:0004076</a> <input type="button" value="add"/> biotin synthase activity (F)                       |  |
| private comment:<br><input type="text" value="Start confidence Low"/>                                                                     |  | public comment:<br><input type="text"/>                                                                                                |  |
| <a href="#">nt_comment</a>                                                                                                                |  | <a href="#">auto_comment</a>                                                                                                           |  |

# Gene Curation Page

## Gene Curation Information

This section contains basic identifying information about the gene and some search and display options.

The **feat\_name** of the gene is listed at the top of the page, this number is called the “gene id” in gene lists in Manatee. The feat\_name is followed in parentheses by the **locus name** (final loci are assigned to genes at the end of a project, once annotation is complete, but they may get temporary loci during the course of the project).

The **blue link** under these names is a link to a file containing the BER search results for this gene (see later slide). There is another link to this page further down the orf info page (will be seen in a later slide).

To the right of the ORF names is a box containing **coordinates, length, and molecular weight**. “end5” is the 5’ coordinate for the beginning of the coding sequence, “end3” is the 3’ coordinate for the end of the coding sequence.

Finally on the extreme right is a box allowing you to move to another ORF info page by typing in the feat\_name or locus in the box and clicking “**new gene**”. One can also change to an orf in a different genome by **changing the database** in the database box, typing in the new orf number and clicking “new gene”.

If you want to reload theGCP, use the “**Reload Page**” link in this section. Do not use the browser’s reload button as this can cause things to be sent to the db in error.

To generate new HMM and BER searches click “**Refresh Searches**” and enter your unix password.

| Shewanella oneidensis MR-1                                                        |  | Gene Curation Page                                                                                 |                                                                     | Home   Logged into [gsp] as mlgwin              |  |
|-----------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|-------------------------------------------------|--|
| <b>GENE CURATION INFORMATION</b>                                                  |  |                                                                                                    |                                                                     |                                                 |  |
| <b>ORF04813 (SO2740)</b><br>▶ View BER Searches<br>asmb_id: 7974<br>▶ Reload Page |  | end5/end3: 2856763 / 2855711<br>gene length: 1053<br>protein length: 350<br>molecular wt: 38790.13 |                                                                     | database: gsp<br>feat_name / locus:<br>New Gene |  |
| Select Display                                                                    |  | Select Function                                                                                    |                                                                     | Refresh Searches                                |  |
| <b>GENE IDENTIFICATION</b> submit   history                                       |  |                                                                                                    |                                                                     |                                                 |  |
| gene name:<br>biotin synthase                                                     |  |                                                                                                    |                                                                     |                                                 |  |
| gene_sym:<br>bioB                                                                 |  |                                                                                                    |                                                                     |                                                 |  |
| EC number(s):<br>2.8.1.6                                                          |  |                                                                                                    | EC GO suggestions:<br>▶ GO:0004076 add biotin synthase activity (F) |                                                 |  |
| private comment:<br>Start confidence Low                                          |  |                                                                                                    | public comment:                                                     |                                                 |  |
| ▶ nt_comment                                                                      |  |                                                                                                    | ▶ auto_comment                                                      |                                                 |  |

# Gene Curation Page

## Gene Identification

Initial information for this section comes from AutoAnnotate. The manual annotation then confirms or changes the information.

**Common name:** the descriptive name given to the protein

**Gene sym:** the gene symbol for the protein (in this case bioB) (we default to E. coli gene symbols when possible and B. subtilis for Gram + specific things)

**EC#:** If the protein is an enzyme, we store the Enzyme Commission number. See later slides for info on ECGO term suggestions.

**private comment:** a field for annotators to note information for later reference by themselves or other annotators. A good place to keep notes.

**public comment:** comments meant to go out with our public accessions.

**auto\_comment:** A link to information from the AutoAnnotate program indicating what information was used to make the preliminary annotation assignments (see next slide).

**nt\_comment:** For non-TIGR comments. This is the place that collaborators can put comments to help the team in annotation.

| Shewanella oneidensis MR-1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |  | Gene Curation Page |  | Home   Logged into [gsp] as mlgwinn |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--------------------|--|-------------------------------------|
| <div> </div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |  |                    |  |                                     |
| <h3>GENE CURATION INFORMATION</h3> <div> <div> <b>ORF04813 (S02740)</b><br/> <a href="#">View BER Searches</a><br/>           asmb_id: 7974<br/> <a href="#">Reload Page</a> </div> <div>           end5/end3: 2856763 / 2855711<br/>           gene length: 1053<br/>           protein length: 350<br/>           molecular wt: 38790.13         </div> <div>           database: <input type="text" value="gsp"/><br/>           feat_name / locus: <input type="text"/><br/> <input type="button" value="New Gene"/> </div> </div> <div> <div> <input type="text" value="Select Display"/> </div> <div> <input type="text" value="Select Function"/> </div> <div> <input type="button" value="Refresh Searches"/> </div> </div> |  |                    |  |                                     |
| <h3>GENE IDENTIFICATION</h3> <div> <a href="#">submit</a>   <a href="#">history</a>    </div> <div> <div> <b>gene name:</b><br/> <input type="text" value="biotin synthase"/> </div> <div> <b>gene_sym:</b><br/> <input type="text" value="bioB"/> </div> <div> <b>EC number(s):</b><br/> <input type="text" value="2.8.1.6"/> </div> <div> <b>private comment:</b><br/> <input type="text" value="Start confidence Low"/> </div> <div> <b>public comment:</b><br/> <input type="text"/> </div> </div> <div> <div> <b>EC GO suggestions:</b><br/> <a href="#">GO:0004076</a> <input type="button" value="add"/> biotin synthase activity (F)         </div> <div> <b>nt_comment</b> </div> <div> <b>auto_comment</b> </div> </div>  |  |                    |  |                                     |

# Gene Curation Page - Auto Comment

Clicking on “auto\_comment” pops up a text box with information on where AutoAnnotate got the information it used for the preliminary annotation.

The screenshot shows a web interface for gene curation. A warning dialog box is open in the center, displaying identification information. The background page is partially visible, showing sections for 'GENE CURATION INFORMATION' and 'GENE IDENTIFICATION'.

**Warning Dialog Box:**

- IDENT INFO:**
  - com\_name = biotin synthetase from TIGR00433
  - gene\_sym = (bioB) from TIGR00433
  - role\_id = 77 from TIGR00433
  - autoannotate = 1
- BEST HMM INFO:**
  - acc: TIGR00433
  - name: biotin synthetase
  - score: 564.1
  - cutoff: 300.00
  - iso\_type: equivalog
  - hit length: full
  - BER hit: NONE

**Background Page Elements:**

- GENE CURATION INFORMATION:**
  - ORF04813 (SO274)
  - [View BER Searches](#) (asmb\_id: 7974)
  - [Reload Page](#)
  - Select Display
- GENE IDENTIFICATION:**
  - gene name: biotin synthase
  - gene\_sym: bioB
  - EC number(s): 2.8.1.6
  - private comment: Start confidence Low
  - nt\_comment
- EC GO suggestions:**
  - [GO:0004076](#) add biotin synthase activity (F)
  - public comment:
  - auto\_comment
- Navigation/Tools:**
  - Back, Forward, Reload, Stop
  - Home, Bookmarks, Welcome to M
  - Search, Print
  - submit, history

# Gene Curation Page - BER Skim and Characterized Match

The characterized match section is where we enter the accession of a match gene whose function has been characterized in the lab (as opposed to having received its name based on sequence similarity.) This is stored as a piece of annotation evidence. This accession will pop into the go with\_ev field in the proper format if you click on “Add to GO Evidence”. (more on GO data later)

The BTAB SKIM section shows the top hits from the BER search file (see Annotation Overview presentation for more information on BER searches). The first column is the accession of the match protein (from various databases), the second is the percent similarity of the match, the third is the length of the match (in nucleotides), the fourth is the name of the match protein and finally, the P score from the BLAST search.

The color of the background for each entry in the skim indicates whether it is in the characterized table and at what confidence level: **green**=high confidence; **red**=automated process; **sky blue**=partial characterization; **olive**=trusted, used when multiple extremely good lines of evidence exist for function but no experiment has been done; **blue-green**=fragment/domain has been characterized; **fuzzy gray**=void, used to indicate that something that was originally thought to be characterized really is not; **gray**=omnium only

Clicking on the **blue accession number** will automatically populate the “Add accession” field in the characterized match section with that accession. Clicking on the **blue names of the proteins** in the skim will take you to a page with just the alignment to that protein.

The blue “**View BER searches**” link at the top of the skim section will take you to a file of all of the pairwise alignments from the BER search (see later slide). The **tree icon** takes you to a phylogenetic tree of the genome protein with the top hits of the skim, the **Belvu icon** takes you to a multiple alignment of the

CHARACTERIZED MATCH

submit

history

(aln) SP:P12996

coords: 7 / 350

score: 42

Pvalue: 7.2e-120

per\_id: 66.0%

per\_sim: 79.7%

[\[Add To GO Evidence\]](#)

Delete accession:

Add accession:

[\[Add To GO Evidence\]](#)

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}  GP13  | 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}  GP19655583lg | 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}  GP19946364lgb  | 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}  GP21  | 6.6e-110 |
| OMNI:NTL02BA0265  | 77.0  | 340    | biotin synthase {Buchnera aphidicola Sg}  GP21623185lgb/AAM6  | 1.4e-109 |
| OMNI:NTL01XF00065 | 79.4  | 309    | biotin synthase {Xylella fastidiosa 9a5c}  GP19104834lgb/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}  GP13425251lgb  | 3.0e-105 |
| OMNI:NTL01BMA0776 | 79.8  | 311    | BIOTIN SYNTHASE {Brucella melitensis 16M}  GP17984969lgb/AAL  | 6.3e-105 |

# Links from the Gene Curation Page - The BER alignment file

This page is accessible by clicking on the “View BER searches” link at the top of the Info page or at the top of the BTAB skim section.

Here you will find multiple pairwise alignments of the genome protein to hits found in the BER search.

In the header of each alignment will be listed the accessions and names for this protein from every database where it is found. These accessions are clickable objects and will take you to the page for the match protein in the database in question.

The background color of the header will be gold if the protein is found in the characterized table with the confidence level indicated by the color of the text for the accession found in this alignment. (This is seen for the SP accession in this alignment.)

Names in Skim are first entry in header, not necessarily the name you want to use, check role notes for TIGR naming standards, check IUBMB EC site for official enzyme names, look in header for SwissProt as a model for the name if previous two guides are not available.

The background color in the Skim may be assigned to an entry in the header different than the one named in the Skim.

Links to info pages for the match protein in the source db.

Shewanella oneidensis M-1 | BER Display For ORF04813 | Home | Logged into [gsp] as mlgwion

Gene Curation Page

66.0/79.7% over 343aa | Escherichia coli

- SP|P12996|Biotin synthase (EC 2.8.1.6) (Biotin synthetase). [Edit characterized](#)
- PIR|J2517|SYEBB biotin synthase (EC 2.8.1.6) bioB [validated] - Escherichia coli (strain K-12) [Insert characterized](#)
- GB|AAC73862.1|P1786992|AE000180 biotin synthesis, sulfur insertion? {Escherichia coli K12;} [Insert characterized](#)

ORF04813( 7 - 350 of 351 aa)  
SP|P12996|BIOB\_ECOLI(4 - 346 of 346) Biotin synthase (EC 2.8.1.6) (Biotin synthetase). {Escherichia coli;} ^|\*PIR|J2517|SYEBB  
%Match = 42.2  
%Identity = 66.0 %Similarity = 79.7  
Matches = 227 Mismatches = 69 Conservative Sub.s = 47  
Gaps = 1 InDels = 3 Frame Shifts = 0  
Primary Frame = 1 [343, 0, 0]

tctctgtgccccacgcacgtgcacacggcggttataggatggtcacacagatgtacagtagattggactcttctagtgcatttc  
gaatagagctccggtgagtcgaataaaccgaagcaaggaagtagcaatcctaaacttataagctctctagtgtcac  
ctaactctaatactctgtacaagcgaggtgtcatgtatataatcaagcaggtataagacgctgggtagggcaga  
-91 -61 -51 -41 -31 -21 -11  
CHQ\*VYGRFPFAPRSLGHCVPKQEVHSHRYKCAKYGRYQSRNVCA\*KL\*ALIE\*SVVNLXHWLALTV\*GLRLS\*P

ataaaaagtattctcgccgtacggaggttgccaaagtttagcaaacgggtgcaggaactttaaaagtcggtattccagctg  
gaacataggctcaatgaagagaatactctcttaatttacagtagaaaaacaataggttctacgcgcgaagaagcagcga  
gataaaagttagggatatttgggaacacatgggtcaactcctagcttcagcccaaggcatgttggttattggtgtcc  
-1 10 20 30 40 50 60 70  
R\*NTKIKGCSMSQLQVRHDKREEIEALFALPNDLLFKAHSTHREYDPNEVQISRLLSIKTGACPECKYCKPQASRYD  
MAHRPRWTLSSQVTELFEPKPLDLLFPAQVVRHDFRQVSTLSIKTGACPECKYCKPQSSRYK  
10 20 30 40 50 60

agcgagcgtgagcagcagcaggggtcttagggtcacagagactcaacagcgagcagataatgataggcagtggggc  
cgtaaagttctactcaacgcacgcgcgtgtgcccgcacaaaatcaataactgtactgtctgtgcaacaatcacgt  
tctaagtcagcagcagcagcaggtgtgctgctgtcgtatgacagcagcagcagcagcagcagcagcagcagcagc  
80 90 100 110 120 130 140 150  
TGLEKERLLAMETVLTARSAAKASRPFMGAAWRNPKDKMPYKQMVQVKALGMECTMLGMLSAEQANLEAGL  
TGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAWKNPHERDMPYLEQMVQVKAMGLEACMTLGLTSESQAQLANAGL  
80 90 100 110 120 130 140

gttacatgactgctgggaacatcactgatacgcgtgaagttggaggaggagagaggttcccgatccccgtgcaaatgag  
aaaaatacccaagaagattccgaagtagctgacgtatgcgtgttgtaaccagcgttaatacatcaactctatttat  
ctccctatgctaccctgacctotaccatcactgcagcagatctcctccgctgctcactaaagttacgtgttggtgacaa  
160 170 180 190 200 210 220 230  
DYNNHNLDSPEYIGDVITRTYQNLDTLSHVRASMKVCSGGIVGMEKATDRAGLLQQLANLPQHPDVPINMLVKV  
DYNNHNLDSPEYIGNITRTYQNLDTLEKVRDAGIKVCSGGIVGLGTVKDRAGLLQQLANLPFPSPVPINMLVKV  
160 170 180 190 200 210 220

ggactgacggtgcccgtgcaaggcattacctggtggcgaaggcagatttgggatattgtatcaacacggaggagtcc  
cgctcaataactattgtctgttctgtgtccgaatgaatactgttgcactttaggtatcccaagaagattgttg  
gtcctaatttatacgtcacccggttagaggggtacaactatgtaggcgtctgcgcttctcagcagcccaatttggcgctc  
240 250 260 270 280 290 300 310  
ACTPFEKLLDLDLEFVRTIAVARILMPLSRVLSAGRENNMSDELOAMCFAGANSIFYGCKILLTPNPEESDMLFRR  
KGTPLANDDDVDAFDFIRTTIAVARIMPTSYVRLSAGRENNMSDELOAMCFAGANSIFYGCKILLTPNPEEDKDLQFRK  
240 250 260 270 280 290 300

cgctcgggggttagggcggtgaggtgctgagtggtgggtgcacagagacgttgtttgactaaacagctgtgtgtaata  
tgtgcaagccctaaactcaacccaaacccataccctacattcctatctgtatggcaatgggagcgctgtct  
gtactggcacttttgagaatatttatatatgtttgaaaatattacatgcgctggacatcaacgagcagcagcagcagc  
320 330 340 350 360 370 380 390  
LGLRPEQGAASIDDEQAVLAKAAAYQDKASQFYDAAL\*PKLIATVKLASLDLCPVKL\*STNPKV\*CG\*SGSARI\*AI  
LGLNPQQTAVLACDNEQQRLRQA-LMTPDDEYYNAAL  
320 330 340

## BER Alignment detail: Boxed Header

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|
| 66.0/79.7% over 343aa                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | <i>Escherichia coli</i> |
| <ul style="list-style-type: none"><li>• <a href="#">SP P12996 </a>Biotin synthase (EC 2.8.1.6) (Biotin synthetase). <a href="#">Edit characterized</a></li><li>• <a href="#">PIR JC2517 </a>SYECBB biotin synthase (EC 2.8.1.6) bioB [validated] - Escherichia coli (strain K-12) <a href="#">Insert characterized</a></li><li>• <a href="#">GB AAC73862.1 </a><a href="#">GP 1786992 </a>AE000180 biotin synthesis, sulfur insertion? {Escherichia coli K12;} <a href="#">Insert characterized</a></li></ul> |                         |

-The background color of this box will be gold if the protein is in the characterized table and grey if it is not.

-The top bar lists the percent identity/similarity and the organism from which the protein comes (if available).

-The bottom section lists all of the accession numbers and names for all the instances of the match protein from the source databases (used in building NIAA for the searches.)

-The accession numbers are links to pages for the match protein in the source databases.

-A particular entry in the list will have colored text (the color corresponding to its characterized status) if that is the accession that is entered into the characterized table - this tells the annotators which link they should follow to find experimental characterization information. Only one accession for the match protein need be in the characterized table for the header to turn gold.

-There are links at the end of each line to enter the accession into the characterized table or to edit an already existing entry in the characterized table.

## BER Alignment detail: alignment header

```
ORF04813( 7 - 350 of 350 aa)  
SP|P12996|BIOB_ECOLI(4 - 346 of 346) Biotin synthase (EC 2.8.1.6)  
%Match = 42.3  
%Identity = 66.0 %Similarity = 79.7  
Matches = 227 Mismatches = 69 Conservative Sub.s = 47  
Gaps = 1 InDels = 3 Frame Shifts = 0  
Primary Frame = 1 [343, 0, 0]
```

-It is most important to look at the range over which the alignment stretches and the percent identity

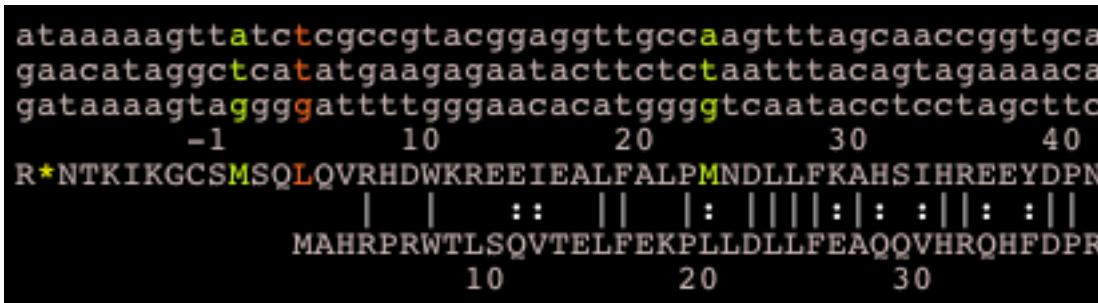
-The top line show the amino acid coordinates over which the match extends for our protein

-The second line shows the amino acid coordinates over which the match extends for the match protein, along with the name and accession of the match protein

-The last line indicates the number of amino acids in the alignment found in each forward frame for the sequence as defined by the coordinates of the gene. The primary frame is the one starting with nucleotide one of the gene. If all is well with the protein, all of the matching amino acids should be in frame 1.

-If there is a frameshift in the alignment (see overview) the phrase "Frame Shifts = #" will flash and indicate how many frameshifts there are.

## BER Alignment detail: alignment of amino acids



-In these alignments the codons of the DNA sequence read down in columns with the corresponding amino acid underneath.

-The numbers refer to amino acid position. Position 1 is the first amino acid of the protein. The first nucleotide of the codon coding for amino acid 1 is nucleotide 1 of the coding sequence. Negative amino acid numbers indicate positions upstream of the predicted start of the protein.

-Vertical lines between amino acids of our protein and the match protein (bottom line) indicate exact matches, dotted lines (colons) indicate similar amino acids.

-Start sites are color coded: ATG is green, GTG is blue, TTG is red/orange

-Stop codons are represented as asterisks in the amino acid sequence. An open reading frame goes from an upstream stop codon to the stop at the end of the protein, while the gene starts at the chosen start codon.

# Swiss-Prot entry - slide #1 - top of page

SwissProt is an incredibly useful database for manual annotation. All of the genes in SwissProt have been manually annotated by an experienced knowledgeable staff. In addition, along with each protein's annotation is stored additional information on references that describe the protein, cross referenced databases in which the protein can be found, motifs which the protein contains, and coordinates of any known features in the protein (and much more.)

**NiceProt View of Swiss-Prot: [P12996](#)**

[Printer-friendly view](#) [Submit update](#) [Quick BlastP search](#)

[\[Entry info\]](#) [\[Name and origin\]](#) [\[References\]](#) [\[Comments\]](#) [\[Cross-references\]](#) [\[Keywords\]](#) [\[Features\]](#) [\[Sequence\]](#) [\[Tools\]](#)

*Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.*

**Entry information**

|                                   |                           |
|-----------------------------------|---------------------------|
| Entry name                        | <b>BIOB_ECOLI</b>         |
| Primary accession number          | <b>P12996</b>             |
| Secondary accession numbers       | None                      |
| Entered in Swiss-Prot in          | Release 13, January 1990  |
| Sequence was last modified in     | Release 35, November 1997 |
| Annotations were last modified in | Release 44, July 2004     |

**Name and origin of the protein**

|              |                                                                                                                                                                                                          |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Protein name | <b>Biotin synthase</b>                                                                                                                                                                                   |
| Synonyms     | <b>EC 2.8.1.6</b><br><b>Biotin synthetase</b>                                                                                                                                                            |
| Gene name    | <b>Name: bioB</b><br>OrderedLocusNames: b0775                                                                                                                                                            |
| From         | <a href="#">Escherichia coli</a> [TaxID: <a href="#">562</a> ]                                                                                                                                           |
| Taxonomy     | <a href="#">Bacteria</a> ; <a href="#">Proteobacteria</a> ; <a href="#">Gammaproteobacteria</a> ; <a href="#">Enterobacteriales</a> ; <a href="#">Enterobacteriaceae</a> ; <a href="#">Escherichia</a> . |

**References**

- [1] SEQUENCE FROM NUCLEIC ACID.  
MEDLINE=89066784;PubMed=3058702 [[NCBI](#), [ExpASY](#), [EBI](#), [Israel](#), [Japan](#)]  
[Otsuka A.J.](#), [Buonocristiani M.R.](#), [Howard P.K.](#), [Flamm J.](#), [Johnson O.](#);  
"The Escherichia coli biotin biosynthetic enzyme sequences predicted from the nucleotide sequence of the bio operon.";  
[J. Biol. Chem. 263:19577-19585\(1988\).](#)
- [2] SEQUENCE FROM NUCLEIC ACID.  
[Pearson B.M.](#), [McKee R.A.](#);  
"Genetic material for expression of biotin synthetase enzymes.";  
Patent number [GB2216530](#), 11-OCT-1989.
- [3] SEQUENCE FROM NUCLEIC ACID.  
**STRAIN=K12 / MG1655**;  
MEDLINE=97426617;PubMed=9278503 [[NCBI](#), [ExpASY](#), [EBI](#), [Israel](#), [Japan](#)]  
[Blattner F.R.](#), [Plunkett G. III](#), [Bloch C.A.](#), [Perna N.T.](#), [Burland V.](#), [Riley M.](#), [Collado-Vides J.](#), [Glasner J.D.](#), [Rode C.K.](#), [Mayhew G.F.](#), [Gregor J.](#), [Davis N.W.](#), [Kirkpatrick H.A.](#),  
[Goeden M.A.](#), [Rose D.J.](#), [Mau B.](#), [Shao Y.](#);  
"The complete genome sequence of Escherichia coli K-12.";  
[Science 277:1453-1474\(1997\).](#)
- [4] CHARACTERIZATION.  
PubMed=8142361 [[NCBI](#), [ExpASY](#), [EBI](#), [Israel](#), [Japan](#)]  
[Sanval I.](#), [Cohen G.](#), [Flint D.H.](#);  
"Biotin synthase: purification, characterization as a [2Fe-2S] cluster protein, and in vitro activity of the Escherichia coli bioB gene product.";  
[Biochemistry 33:3625-3631\(1994\).](#)
- [5] MUTAGENESIS OF CYSTEINE RESIDUES.  
MEDLINE=21547190;PubMed=11686025 [[NCBI](#), [ExpASY](#), [EBI](#), [Israel](#), [Japan](#)]

accession and  
version  
information

name, EC#  
gene\_symbol  
taxonomy

references with  
links to  
abstracts (click  
on NCBI to see  
a PubMed  
abstract of the  
paper)

Link to Enzyme Commission page  
(see later slide)

# Swiss-Prot entry - slide #2 - middle of page

useful  
functional  
information

links to  
other dbs  
where the  
protein is  
found or to  
motif  
clusters or  
protein  
families  
which this  
protein is a  
member of

| Comments                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>• <b>CATALYTIC ACTIVITY:</b> Dethiobiotin + sulfur = biotin.</li> <li>• <b>COFACTOR:</b> Binds a 4Fe-4S cluster coordinated with 3 cysteines and an exchangeable S-adenosyl-L-methionine, and a 2Fe-2S cluster coordinated with 3 cysteines and 1 arginine.</li> <li>• <b>PATHWAY:</b> Biotin biosynthesis; last step.</li> <li>• <b>SUBUNIT:</b> Homodimer.</li> <li>• <b>SIMILARITY:</b> Belongs to the biotin and lipoic acid synthetases family.</li> </ul> |
| Copyright                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcements/">http://www.isb-sib.ch/announcements/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ) |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Cross-references                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| EMBL                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | J04423; AAA23515.1; -. <a href="#">[EMBL]</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> / <a href="#">CoDingSequence</a><br>A11530; CAA00965.1; -. <a href="#">[EMBL]</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> / <a href="#">CoDingSequence</a><br>AE000180; AAC73862.1; -. <a href="#">[EMBL]</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a> / <a href="#">CoDingSequence</a>                                                                                                |
| PIR                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <a href="#">JC2517</a> ; SYECBB.                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| PDB                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 1R30; 13-JAN-04. <a href="#">[ExpASY]</a> / <a href="#">RCSB</a> / <a href="#">EBI</a>                                                                                                                                                                                                                                                                                                                                                                                                                 |
| ECO2DBASE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <a href="#">E038.6</a> ; 6TH EDITION.                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| EchoBASE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <a href="#">EB0116</a> ; -.                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| EcoGene                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <a href="#">EG10118</a> ; bioB.                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| EcoCyc                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <a href="#">EG10118</a> ; bioB.                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| CMR                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <a href="#">P12996</a> ; b0775.                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| InterPro                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <a href="#">IPR010722</a> ; BATS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <a href="#">IPR002684</a> ; Biotin_synth.                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <a href="#">IPR006638</a> ; Elp3/MiaB/NifB.                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <a href="#">IPR007197</a> ; Radical_SAM.                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <a href="#">Graphical view of domain structure.</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Pfam                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <a href="#">PF06968</a> ; BATS; 1.                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <a href="#">PF04055</a> ; Radical_SAM; 1.                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <a href="#">Pfam graphical view of domain structure.</a>                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| SMART                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | <a href="#">SM00729</a> ; Elp3; 1.                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| TIGRFAMs                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <a href="#">TIGR00433</a> ; bioB; 1.                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| ProDom                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <a href="#">[Domain structure / List of seq. sharing at least 1 domain]</a>                                                                                                                                                                                                                                                                                                                                                                                                                            |
| HOBACGEN                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <a href="#">[Family / Alignment / Tree]</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| BLOCKS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <a href="#">P12996</a> .                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| ProtoNet                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <a href="#">P12996</a> .                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| ProtoMap                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <a href="#">P12996</a> .                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| PRESAGE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <a href="#">P12996</a> .                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| DIP                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <a href="#">P12996</a> .                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| ModBase                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <a href="#">P12996</a> .                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| SMR                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <a href="#">P12996</a> ; 550A7899A2DF6082.                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| SWISS-2DPAGE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <a href="#">Get region on 2D PAGE.</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| UniRef                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <a href="#">View cluster of proteins with at least 50% / 90% identity.</a>                                                                                                                                                                                                                                                                                                                                                                                                                             |

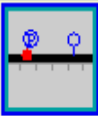
# Swiss-Prot entry - slide #3 - bottom of page

## keywords and sequence features with coordinates

Keywords

2Fe-2S; 3D-structure; 4Fe-4S; Biotin biosynthesis; Complete proteome; Iron-sulfur; Transferase.

Features


[Feature table viewer](#)

| Key      | From                | To                  | Length | Description                          |
|----------|---------------------|---------------------|--------|--------------------------------------|
| METAL    | <a href="#">53</a>  | <a href="#">53</a>  |        | Iron-sulfur 1 (4Fe-4S).              |
| METAL    | <a href="#">57</a>  | <a href="#">57</a>  |        | Iron-sulfur 1 (4Fe-4S).              |
| METAL    | <a href="#">60</a>  | <a href="#">60</a>  |        | Iron-sulfur 1 (4Fe-4S).              |
| METAL    | <a href="#">97</a>  | <a href="#">97</a>  |        | Iron-sulfur 2 (2Fe-2S).              |
| METAL    | <a href="#">128</a> | <a href="#">128</a> |        | Iron-sulfur 2 (2Fe-2S).              |
| METAL    | <a href="#">188</a> | <a href="#">188</a> |        | Iron-sulfur 2 (2Fe-2S).              |
| METAL    | <a href="#">260</a> | <a href="#">260</a> |        | Iron-sulfur 2 (2Fe-2S).              |
| CONFLICT | <a href="#">63</a>  | <a href="#">63</a>  |        | S -> T (in Ref. <a href="#">1</a> ). |

sequence features

Sequence information

Length: 346 AA

Molecular weight: 38648 Da

CRC64: 550A7899A2DF6082 [This is a checksum on the sequence]

|            |            |            |            |            |            |
|------------|------------|------------|------------|------------|------------|
| 10         | 20         | 30         | 40         | 50         | 60         |
|            |            |            |            |            |            |
| MAHDPNWTLS | QVTELEIKPL | LDLLFIAQOV | HRQHEDPRQV | QVSTLLSIKT | GACPEDCKYC |
| 70         | 80         | 90         | 100        | 110        | 120        |
|            |            |            |            |            |            |
| PQSSRYKTGL | EAERLMEVIQ | VLESARKAKA | AGSTRICMGA | AWKNPHEEDM | PYLEQWQGV  |
| 130        | 140        | 150        | 160        | 170        | 180        |
|            |            |            |            |            |            |
| KAMGLEACMT | LGTLSESQAR | RLANAGLDYY | NANLDTSEEF | YGNIIITTTY | QERLDTLEKY |
| 190        | 200        | 210        | 220        | 230        | 240        |
|            |            |            |            |            |            |
| RDAGIKVCSG | GIVGLGITYK | DRAGLLQLA  | NLPTPEESVP | INMLVKVKG  | PLADNDVDA  |
| 250        | 260        | 270        | 280        | 290        | 300        |
|            |            |            |            |            |            |
| FDIIRTIARA | RIMPTSYVR  | LSAGRIQME  | QTQAMCMAG  | ANSIFYGCKL | LTTPNPEEDK |
| 310        | 320        | 330        | 340        |            |            |
|            |            |            |            |            |            |
| DLQLEFKLGL | NPOQTAVLAG | DNEQQRLIQ  | ALPTPTDEY  | YNAAAL     |            |

P12996 in [FASTA format](#)

sequence features

# View of EC number info page from Swiss Institute of Bioinformatics site

## NiceZyme View of ENZYME: EC [2.8.1.6](#)

|                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Official Name</b>                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Biotin synthase.                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Alternative Name(s)</b>                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Biotin synthetase.                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Reaction catalysed</b>                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Dethiobiotin<br>+ sulfur<br>=><br>biotin                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Cofactor(s)</b>                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Iron-sulfur.                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Comments</b>                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <ul style="list-style-type: none"><li>The sulfur donor has been unidentified to date - it is not elemental sulfur or an iron-sulfur cluster.</li></ul> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Cross-references</b>                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| BRENDA                                                                                                                                                 | <a href="#">2.8.1.6</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| EMP/PUMA                                                                                                                                               | <a href="#">2.8.1.6</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| WIT                                                                                                                                                    | <a href="#">2.8.1.6</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Kyoto University LIGAND chemical database                                                                                                              | <a href="#">2.8.1.6</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| IUBMB Enzyme Nomenclature                                                                                                                              | <a href="#">2.8.1.6</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| IntEnz                                                                                                                                                 | <a href="#">2.8.1.6</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| MEDLINE                                                                                                                                                | <a href="#">Find literature relating to 2.8.1.6</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Swiss-Prot                                                                                                                                             | <a href="#">P54967</a> , <a href="#">BIOB_ARATH</a> ; <a href="#">P19206</a> , <a href="#">BIOB_BACSH</a> ; <a href="#">P53557</a> , <a href="#">BIOB_BACSU</a> ;<br><a href="#">P57378</a> , <a href="#">BIOB_BUCAI</a> ; <a href="#">Q8K9P1</a> , <a href="#">BIOB_BUCAP</a> ; <a href="#">Q89AK5</a> , <a href="#">BIOB_BUCBP</a> ;<br><a href="#">P12997</a> , <a href="#">BIOB_CITFR</a> ; <a href="#">P46396</a> , <a href="#">BIOB_CORGL</a> ; <a href="#">P12996</a> , <a href="#">BIOB_ICOLI</a> ;<br><a href="#">Q47862</a> , <a href="#">BIOB_ERWHE</a> ; <a href="#">P44987</a> , <a href="#">BIOB_HAEIN</a> ; <a href="#">Q92JK8</a> , <a href="#">BIOB_HELPJ</a> ;<br><a href="#">Q25956</a> , <a href="#">BIOB_HELPY</a> ; <a href="#">Q58692</a> , <a href="#">BIOB_METVA</a> ; <a href="#">P94966</a> , <a href="#">BIOB_METSK</a> ;<br><a href="#">P46715</a> , <a href="#">BIOB_MYCLE</a> ; <a href="#">Q06601</a> , <a href="#">BIOB_MYCTU</a> ; <a href="#">P12678</a> , <a href="#">BIOB_SALTY</a> ;<br><a href="#">Q59778</a> , <a href="#">BIOB_SCHPO</a> ; <a href="#">P36569</a> , <a href="#">BIOB_SERMA</a> ; <a href="#">P73538</a> , <a href="#">BIOB_SYNY3</a> ;<br><a href="#">P32451</a> , <a href="#">BIOB_YEAST</a> ; |

[View entry in original ENZYME format](#)

All Swiss-Prot entries referenced in this entry, with possibility to download in different formats, align etc.

Link to official Enzyme Commission site

# View of information page for an EC number at IUBMB site

The Enzyme Commission (EC) is part of the IUBMB and is charged with maintaining the database of enzyme classifications. In the EC system, each reaction is assigned a 4 part accession number with each part consisting of an integer, where the numbers are separated by periods. As one moves from the first number to the second to the third to the fourth the nature of the reaction becomes more specific. For example: EC2.-.- = “transferase”, 2.8.- = “transferase, transferring sulfur-containing groups”, 2.8.1.- = “sulfurtransferases”, and finally 2.8.1.6 = “biotin synthase” (a specific sulfurtransferase, which is a specific class of transferases that transfer sulfur-containing groups). One can see the breakdown of all of the classes within each EC first number (they only go up to 6) by clicking on the home page for each number (see below).

IUBMB Enzyme Nomenclature

## EC 2.8.1.6

**Common name:** biotin synthase

**Reaction:** dethiobiotin + sulfur = biotin

**Systematic name:** dethiobiotin:sulfur sulfurtransferase

**Comments:** an iron-sulfur enzyme. The sulfur donor has been unidentified to date - it is not elemental sulfur or an iron-sulfur cluster.

**Links to other databases:** [BRENDA](#), [EXPASY](#), [KEGG](#), [ERGO](#), [PDB](#), CAS registry number: 80146-93-6 (204794-88-7, 179608-56-1, 209603-31-6, 153554-27-9, 174764-24-0 and 215108-34-2)

**References:**

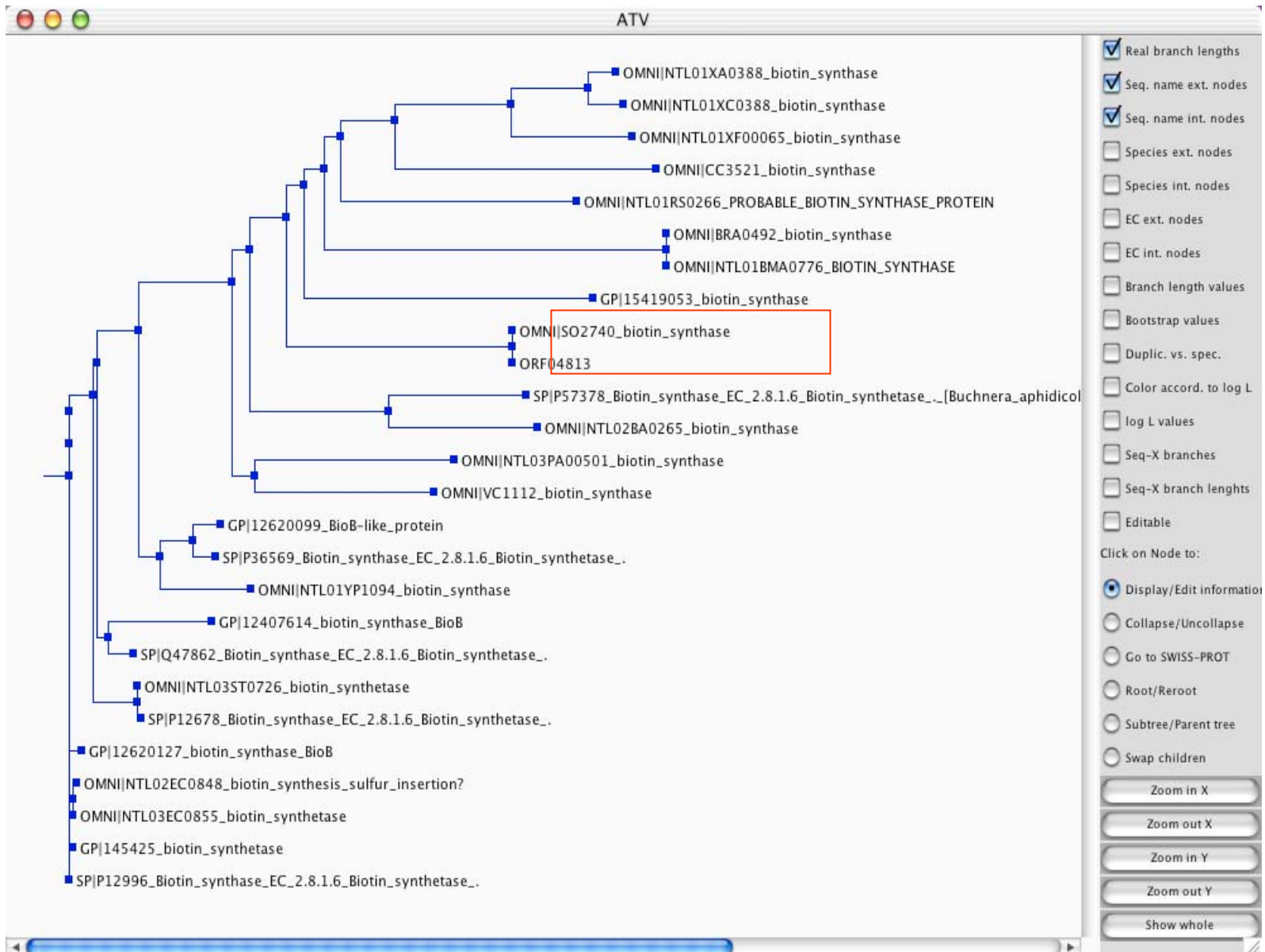
1. Shiuan, D., Campbell, A. Transcriptional regulation and gene arrangement of *Escherichia coli*, *Citrobacter freundii* and *Salmonella typhimurium* biotin operons. *Gene* 67 (1988) 203-211. [Medline UI: [89006280](#)]
2. Zhang, S., Sanyal, I., Bulboaca, G.H., Rich, A., Flint, D.H. The gene for biotin synthase from *Saccharomyces cerevisiae*: cloning, sequencing, and complementation of *Escherichia coli* strains lacking biotin synthase. *Arch. Biochem. Biophys.* 309 (1994) 29-35. [Medline UI: [94161552](#)]

[EC 2.8.1.6 created 1999]

---

Return to [EC 2.8.1 home page](#)  
Return to [EC 2.8 home page](#)  
Return to [EC 2 home page](#) → Click here to see all the classifications within EC #2 (the transferases).  
Return to [Enzymes home page](#)  
Return to [IUBMB Biochemical Nomenclature home page](#)

# Links from the Gene Curation Page - Tree (may not work on laptops)



## Links from the Gene Curation Page - BER multiple alignment

(will not work on laptops)

| File                           |   | Edit                                               | Colour                                     | Sort                                           | Picked:                          |
|--------------------------------|---|----------------------------------------------------|--------------------------------------------|------------------------------------------------|----------------------------------|
| (26x440)                       |   | -20-----30-----40-----50-----60-----70-----80----- |                                            |                                                |                                  |
| OMNI NTL01XA0388               | 1 | 344                                                | .MSVV....LRHDWRKELQALFDL                   | PFPELLHRAASVHRAHFDPAQVQVSTLLSVKTGGCPEDCAYCP    |                                  |
| OMNI NTL01XC0388               | 1 | 344                                                | .MSVV....VRHDWRKELHALFAL                   | PFPELLHRAASVHRAHFDPAEVQVSTLLSVKTGGCPEDCAYCP    |                                  |
| OMNI NTL01RS0266               | 1 | 360                                                | TPGQS....PNARWSREAIEALFAL                  | PFNDLLFQAQQVHRAHFDANAVQLSTLLSIKTGGCPEDCSYCP    |                                  |
| OMNI NTL03PA00501              | 1 | 352                                                | TASVA....TRHDWSLAEVKALFEQPFNDLLFQAQT       | VHRAHFDPNRVQVSTLLSIKTGACPEDCKYCP               |                                  |
| GPI5921547 emb CAB56476.1 AJ2  | 1 | 254                                                | TDACA....TRHDWSLAEVKALFEQPFNDLLFQAQT       | VHRQHFDANRVQVSTLLSIKTGACPEDCKYCP               |                                  |
| ORF06889                       | 1 | 352                                                | STTAT....LRHDWTLAEVRALFVQPFNDLLFQAQT       | LHRAHFDANRVQVSTLLSIKTGACPEDCKYCP               |                                  |
| OMNI VC1112                    | 1 | 350                                                | ...ME....VRHNWTVAEVKALLDKPFMDLLFEAQQVHRL   | HHPHNHVQVSTLLSIKTGACPEDCKYCP                   |                                  |
| OMNI NTL03EC0855               | 1 | 346                                                | ...MA....HRPRWTLSQVTELF                    | FEKPLDLLFEAQQVHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP  |                                  |
| OMNI NTL02EC0848               | 1 | 346                                                | ...MA....HRPRWTLSQVTELF                    | FEKPLDLLFEAQQVHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP  |                                  |
| SPIP12996 BIOB_ECOLI           | 1 | 346                                                | ...MA....HRPRWTLSQVTELF                    | FEKPLDLLFEAQQVHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP  |                                  |
| GPI145425 gb AAA23515.1 J0442  | 1 | 346                                                | ...MA....HRPRWTLSQVTELF                    | FEKPLDLLFEAQQVHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP  |                                  |
| GPI12620127 gb AAG60579.1 AF25 | 1 | 346                                                | ...MA....HRPRWTLSQVTELF                    | FEKPLDLLFEAQQVHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP  |                                  |
| OMNI NTL03ST0726               | 1 | 346                                                | ...MA....RHPRWTLSQVTELF                    | FEKPLLELLFEAQQIHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP |                                  |
| SPIP12678 BIOB_SALTY           | 1 | 346                                                | ...MA....RHPRWTLSQVTELF                    | FEKPLLELLFEAQQIHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP |                                  |
| SPIQ47862 BIOB_ERWHE           | 1 | 346                                                | ...MA....HHARWTLSQVTELF                    | FDKPLDLLFEAQTIRQHFDPRQVQVSTLLSIKTGACPEDCKYCP   |                                  |
| GPI12407614 gb AAG53589.1 AF24 | 1 | 346                                                | ...MA....HQRWTLSQVTALFEKPLLELLFEAQQIHRQHFD | PQQIQVSTLLSIKTGACPEDCKYCP                      |                                  |
| SPIP36569 BIOB_SERMA           | 1 | 346                                                | ..MMA....DRIHWTVGQAQALFDKPLLELLFEAQT       | VHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP               |                                  |
| GPI12620099 gb AAG60559.1 AF2  | 1 | 346                                                | ..MMA....DRIHWTVGLAQT                      | LDKPLLELLFEAQT                                 | VHRQHFDPRQVQVSTLLSIKTGACPEDCKYCP |

# Gene Curation page - HMM hits scoring above noise

(Text describing the features of the HMM section is boxed in the same color as each feature.)

The blue id numbers for each HMM link to an info page for that HMM.

Key information is the isology type and the “total” and “cutoff” scores.

The “Add To GO Evidence” link automatically fills the HMM information into the “with” field in the GO term entry box.

GO terms assigned to each HMM are listed under the HMM (if any). Clicking on the “Add” button here adds not only the GO term id, but also the HMM evidence.

The “Add To Annotation” link will automatically copy the annotation from the HMM to the protein.

Click to see hits below noise

**HMM** [submit](#) | [all hmms](#) | [info](#)

**TIGR00433** **biotin synthase** gene\_sym: **bioB** ec#: **2.8.1.6** role\_id: **77** [\[ Add To Annotation \]](#)

Isology: **equivalog**  
Total score: **564.1** Trusted cutoff: **300.00** Gathering cutoff: **300.00** Noise cutoff: **50.00** Total expect: **1.5e-166**  
Trusted cutoff2: **300.00** Gathering cutoff2: **300.00** Noise cutoff2: **50.00**

| View Alignment             | Coords | HMM Coords  | Score | Expect   | Curation                            |
|----------------------------|--------|-------------|-------|----------|-------------------------------------|
| <a href="#">align page</a> | 18-313 | 1-350 / 350 | 564.1 | 1.5e-166 | <input checked="" type="checkbox"/> |

[GO:0004076](#) [add](#) **biotin synthase activity (F)**

[GO:0009102](#) [add](#) **biotin biosynthesis (P)**

**Genome Properties** [Run Rules.spl...](#) [add GO evidence](#) [\[ GO \]](#) This section described on later slide

| state | property name              |
|-------|----------------------------|
| YES   | <b>biotin biosynthesis</b> |

**PF06968: Biotin and Thiamin Synthesis associated domain** gene\_sym: **none** ec#: **none** role\_id: **none** [\[ Add To Annotation \]](#)

Isology: **domain**  
Total score: **181.7** Trusted cutoff: **43.70** Gathering cutoff: **25.00** Noise cutoff: **19.10** Total expect: **9.8e-52**  
Trusted cutoff2: **43.70** Gathering cutoff2: **25.00** Noise cutoff2: **19.10**

| View Alignment             | Coords  | HMM Coords  | Score | Expect  | Curation                            |
|----------------------------|---------|-------------|-------|---------|-------------------------------------|
| <a href="#">align page</a> | 223-315 | 1-115 / 115 | 181.7 | 9.8e-52 | <input checked="" type="checkbox"/> |

[No HMM-GO Suggestions To Display.](#)

**PF04055: radical SAM domain protein** gene\_sym: **none** ec#: **none** role\_id: **703** [\[ Add To Annotation \]](#)

Isology: **domain**  
Total score: **82.7** Trusted cutoff: **8.80** Gathering cutoff: **8.40** Noise cutoff: **8.30** Total expect: **6.1e-22**  
Trusted cutoff2: **8.80** Gathering cutoff2: **8.40** Noise cutoff2: **8.30**

| View Alignment             | Coords | HMM Coords  | Score | Expect  | Curation                            |
|----------------------------|--------|-------------|-------|---------|-------------------------------------|
| <a href="#">align page</a> | 50-212 | 1-163 / 163 | 82.7  | 6.1e-22 | <input checked="" type="checkbox"/> |

[GO:0003824](#) [add](#) **catalytic activity (F)**

[GO:0008152](#) [add](#) **metabolism (P)**

**HMM report page** - to get to this page click on an HMM accession number almost anywhere in Manatee

At the top is information about the HMM including HMM name, associated annotation (gene symbol, EC#, TIGR role, etc.) and comments from the authors.

Below is a list of all genes in the organism which hit the HMM and the scores they received. The row with the gold background is the protein of interest. Rows with a green background have scores below the trusted cutoff, rows with a purple background have scores below the noise cutoff.

Shewanella oneidensis MR-1

TIGR00433 HMM Report for ORF04813

[Home](#) | [Logged into \[gsp\] as mlgwinn](#)

This page displays information about a specific HMM accession as it relates to the ORF being annotated. General information about the model is presented, as well as an alignment of the model to the ORF and a list of all hits of this model to the genome. The user can follow links to more information about the model and other proteins that the model hits.

|                    |                                                                                              |                |                     |               |          |
|--------------------|----------------------------------------------------------------------------------------------|----------------|---------------------|---------------|----------|
| accession and name | TIGR00433: biotin synthase                                                                   |                |                     |               |          |
| expanded name      | biotin synthetase                                                                            |                |                     |               |          |
| gene symbol        | bioB                                                                                         | EC number      | 2.8.1.6             | HMM length    | 350      |
| model type         | equivalog                                                                                    | trusted cutoff | 300.00              | noise cutoff  | 50.00    |
| author             | Loftus BJ                                                                                    | created        | 04/20/99            | last modified | 09/23/03 |
| related accession  | IPR002684                                                                                    | accession type | InterPro assignment |               |          |
| role category      | 77: Biosynthesis of cofactors, prosthetic groups, and carriers, Biotin                       |                |                     |               |          |
| gene ontology      | GO:0004076 (function): biotin synthase activity<br>GO:0009102 (process): biotin biosynthesis |                |                     |               |          |
| comment            | Catalyzes the last step of the biotin biosynthesis pathway.                                  |                |                     |               |          |
| private comment    |                                                                                              |                |                     |               |          |

Edit HMM Annotation

HMM Inter Link Editor

All DB Hits to TIGR00433

color key

Protein of Interest.

Scores below trusted cutoff ( < 300.00).

Scores below noise cutoff ( < 50.00).

| feat_name | role_id | EC number | gene region | HMM region | score  | gene name                      |
|-----------|---------|-----------|-------------|------------|--------|--------------------------------|
| ORF04813  | 77      | 2.8.1.6   | 18-313      | 1-350      | 564.1  | biotin synthase                |
| ORF03390  | 157     | -         | 34-331      | 1-350      | -168.2 | biotin synthase family protein |
| ORF01034  | 80      | -         | 76-296      | 1-350      | -178.3 | lipoic acid synthetase         |
| ORF03392  | 162     | -         | 62-370      | 1-350      | -187.3 | thiH protein, putative         |

# Genome Properties - linked from the Gene Curation Page in the HMM section

If an HMM is part of a genome property, there will be a link here and an indication of the state of the property - in this case “YES” indicating that the organism has an intact biotin biosynthesis pathway. Clicking on the name of the property takes one to a property report page.

If you want to use the Genome Property as evidence for GO annotation, click the “GO” link under the “add GO evidence” section. (more on GO data later)

HMM

[submit](#) | [all hmms](#) |

▶ **TIGR00433: biotin synthase**gene\_sym: **bioB**ec#: **2.8.1.6**role\_id: **77**[\[Add To Annotation\]](#)

Isology: **equivalog**

Total score: **564.1**

Trusted cutoff: **300.00**

Gathering cutoff: **300.00**

Noise cutoff: **50.00**

Total expect: **1.5e-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](#)

10-313

1-350 / 350

564.1

1.5e-166

☒

▶ **GO:0004076**

add

**biotin synthase activity (F)**

▶ **GO:0009102**

add

**biotin biosynthesis (P)**

▶ **Genome Properties**

[Run Rules.spl...](#)

state

property name

add GO evidence

YES

**biotin biosynthesis**

[\[GO\]](#)

The “Run Rules.spl” link

## Genome Property info page (part 1): biotin biosynthesis

This has general information about the property, GO terms assigned to the property, and a place for curators to put comments regarding this property in this organism.

| Property Definition     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                  |                    |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|--------------------|
| <b>property:</b>        | biotin biosynthesis                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <b>state:</b>    | YES                |
| <b>property type:</b>   | PATHWAY                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | <b>value:</b>    | 1                  |
| <b>role id:</b>         | 77                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | <b>assignby:</b> | HYBRID             |
| <b>GO ids:</b>          | <a href="#">GO:0006355</a> <input type="button" value="add"/> regulation of transcription, DNA-dependent<br><a href="#">GO:0009102</a> <input type="button" value="add"/> biotin biosynthesis                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <b>date:</b>     | Mar 10 2004 3:51PM |
| <b>description:</b>     | Biotin is an essential cofactor for many carboxylation (addition of CO <sub>2</sub> ) reactions. This property reflects biosynthesis from pimeloyl-CoA. The source of pimeloyl-CoA may vary. BioF (EC 2.3.1.47, 8-amino-7-oxononanoate synthase, also called 7-keto-8-aminopelargonic acid synthetase) converts pimeloyl-CoA to 8-amino-7-oxononanoate. BioA (EC 2.6.1.62, adenosylmethionine-8-amino-7-oxononanoate aminotransferase) converts the product to 7,8-diaminononanoate, from which BioD (EC 6.3.3.3, dethiobiotin synthase) makes dethiobiotin. BioB (EC 2.8.1.6, biotin synthase) then makes biotin itself. Enzymes such as BioH involved in pimeloyl-CoA biosynthesis typically receive biotin-related annotations but may also appear in genomes in which biotin is not synthesized and pimeloyl-CoA is used for something else. |                  |                    |
| <b>auto_comment:</b>    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                  |                    |
| <b>curator comment:</b> | <div> <input type="text"/> <input type="button" value="update"/> </div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                  |                    |

## Genome Property info page (part 2): biotin biosynthesis

| Property Steps                         |   |    |                                                             |                                                            |                           |            |
|----------------------------------------|---|----|-------------------------------------------------------------|------------------------------------------------------------|---------------------------|------------|
| REQUIRED                               |   |    | 8-amino-7-oxo-nonanoate synthase (2)                        |                                                            |                           |            |
| A                                      | C | GC | gene id                                                     | gene name                                                  | evidence                  | role id OP |
|                                        |   |    | ORF04814                                                    | 8-amino-7-oxononanoate synthase                            | GENE_CLUSTER<br>TIGR00858 | 77         |
| REQUIRED                               |   |    | adenosyl methionine 8-amino-7-oxononanoate transaminase (3) |                                                            |                           |            |
| A                                      | C | GC | gene id                                                     | gene name                                                  | evidence                  | role id OP |
|                                        |   |    | ORF04812                                                    | adenosylmethionine-8-amino-7-oxononanoate aminotransferase | TIGR00508                 | 77         |
| REQUIRED                               |   |    | dethiobiotin synthase (4)                                   |                                                            |                           |            |
| A                                      | C | GC | gene id                                                     | gene name                                                  | evidence                  | role id OP |
|                                        |   |    | ORF04817                                                    | dethiobiotin synthase                                      | TIGR00347                 | 77         |
| REQUIRED                               |   |    | biotin synthase (5)                                         |                                                            |                           |            |
| A                                      | C | GC | gene id                                                     | gene name                                                  | evidence                  | role id OP |
|                                        |   |    | ORF04813                                                    | biotin synthase                                            | TIGR00433                 | 77         |
| NOT REQUIRED                           |   |    | BioC (bioC)                                                 |                                                            |                           |            |
| A                                      | C | GC | gene id                                                     | gene name                                                  | evidence                  | role id OP |
|                                        |   |    | ORF04816                                                    | biotin synthesis protein BioC                              | TIGR02072                 | 77         |
| NOT REQUIRED                           |   |    | bioH protein (bioH)                                         |                                                            |                           |            |
| A                                      | C | GC | gene id                                                     | gene name                                                  | evidence                  | role id OP |
|                                        |   |    | ORF02552                                                    | bioH protein                                               | TIGR01738                 | 77         |
| NOT REQUIRED                           |   |    | biotin repressor (represso)                                 |                                                            |                           |            |
| A                                      | C | GC | gene id                                                     | gene name                                                  | evidence                  | role id OP |
| THIS COMPONENT HAS NOT BEEN IDENTIFIED |   |    |                                                             |                                                            |                           |            |

This section of the page shows the steps for the property, which steps are required and which steps are not, and the genes from the genome that have been identified for each step.

One can link to the GCP for each gene or to the HMM info page for the HMMs named by clicking on the gene id or HMM accession, respectively.

## Genome Property info page (part 3): biotin biosynthesis

This section has reference information and a graphic showing the cluster of genes in the organism involved in the property. One can click on the arrows in the graphic to get a GCP for that gene.

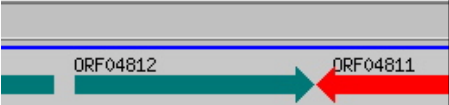
| Property References |                                                           |
|---------------------|-----------------------------------------------------------|
| accession:          | <a href="#">PMID:9847135</a>                              |
| title:              | KEGG: Kyoto Encyclopedia of Genes and Genomes.            |
| authors:            | Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. |
| publication:        | Nucleic Acids Res. 1999 Jan 1;27(1):20-34.                |

| Literature References |  |
|-----------------------|--|
|-----------------------|--|

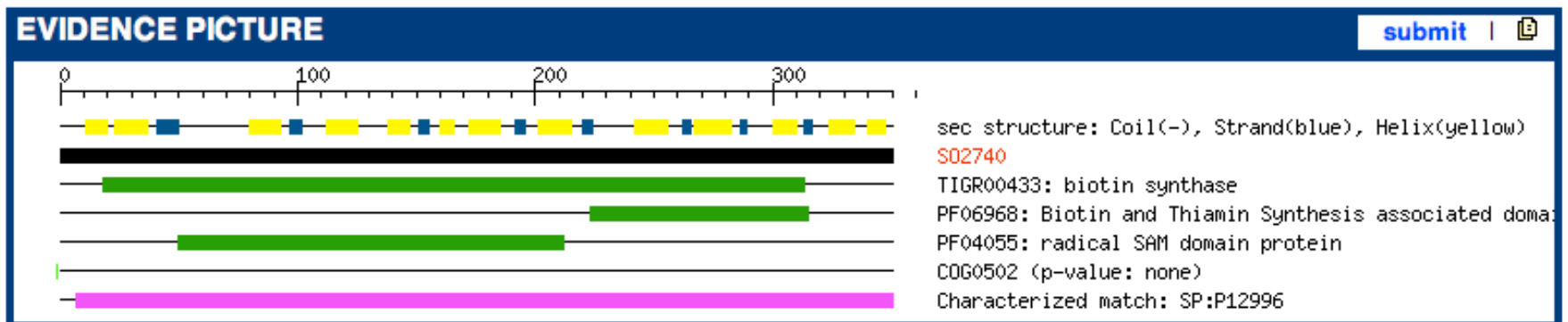
  

| Web References |                                         |
|----------------|-----------------------------------------|
| title:         | <a href="#">KEGG: Biotin Metabolism</a> |

# Gene Curation Page - Evidence Picture - ORF04813

All of the evidence stored for an ORF is displayed in this graphic. The black bar represents the ORF in question. Green bars represent HMMs which hit the ORF above trusted cutoff. Green HMM bars indicate above trusted score, orange indicates above noise but below trusted, red indicates below noise and is generally not shown unless an annotator has decided that the HMM should be included as evidence by toggling the curation box. The pink bar represents the characterized match to this ORF. Characterized matches are shown in different colors that at this time have no meaning. Also shown here is a secondary structure prediction (not run on all genomes). Clicking on the colored bars in the graphic opens windows with additional information on that piece of evidence. To get additional cog info, you must click on the very skinny bar all the way to the left of the cog row. The evidence picture for ORF04813 does not contain all of the possible evidence types, so later slides will show some evidence pictures from other genes.



***Shewanella oneidensis* MR-1 | PSIPRED Prediction Results - ORF04813**

## Overview

### PSIPRED Prediction Alignment

Key:  helix — coil  strand

**The biotin synthase does not have all of the evidence types that are possible, therefore, the following screen shots will show some evidence pictures from other genes displaying additional evidence types.**

**Following the evidence pictures will be the evidence detail pages linked to from the evidence pictures.**

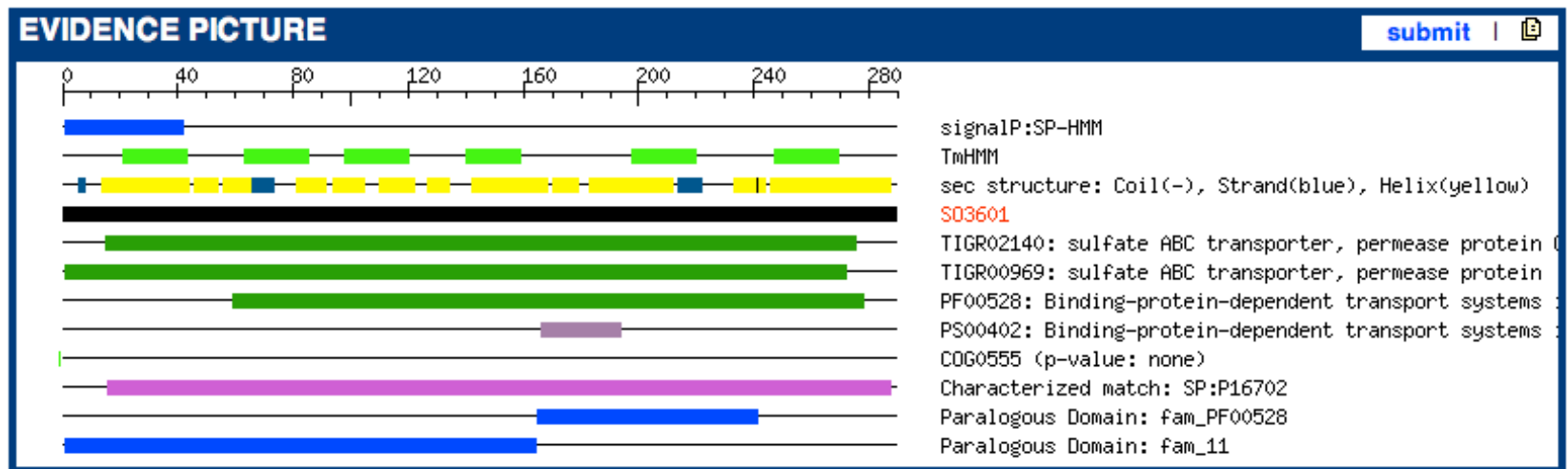
**After all of the evidence types have been represented, the tutorial will resume with ORF04813.**

# Gene Curation Page - Evidence Picture (ORF03779)

Additional evidence types shown here are:

TmHMM - an HMM specific for transmembrane regions, built by the Center for Biological Sequence Analysis, Denmark

Paralogous Family membership - if a protein is a member of a paralogous family it will be represented with a blue bar, clicking on the bar takes you to a page listing all the family members. Paralogous families are built from searching the protein set for a genome against itself. First families are built according to shared hits to HMMs, then regions not matching HMMs are searched against each other to find additional families. The families corresponding to HMMs are given names with the HMM accession number, others are given numbers.



NOTE: this display is from ORF03779

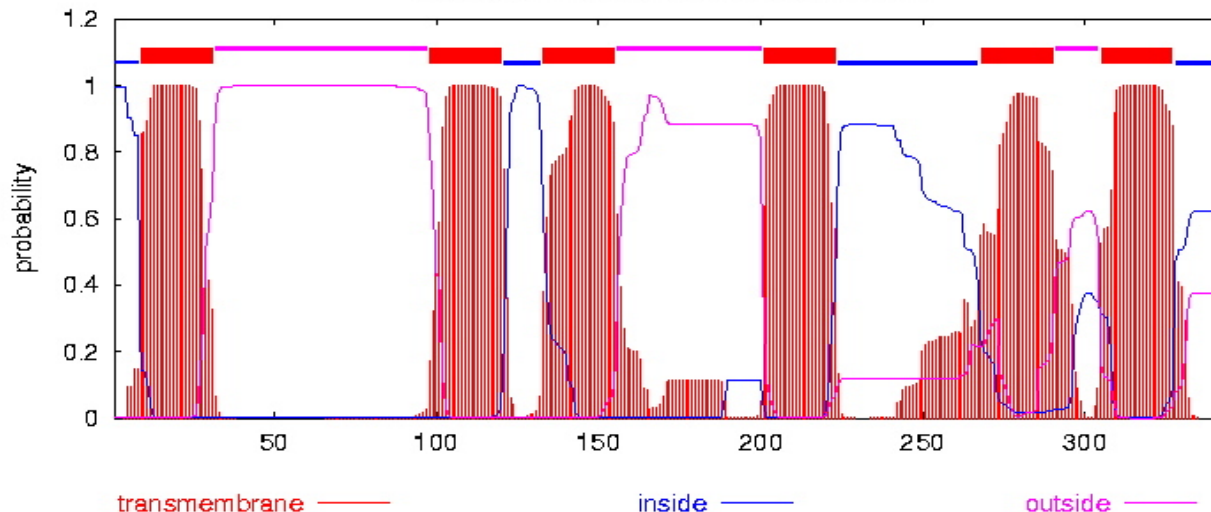
NOTE: this display is for ORF03779

## TMHMM result

[HELP](#) with output formats

```
# Sequence Length: 343
# Sequence Number of predicted TMHs: 6
# Sequence Exp number of AAs in TMHs: 139.48261
# Sequence Exp number, first 60 AAs: 20.99155
# Sequence Total prob of N-in: 0.99734
# Sequence POSSIBLE N-term signal sequence
Sequence      TMHMM2.0      inside      1      8
Sequence      TMHMM2.0      TMhelix     9      31
Sequence      TMHMM2.0      outside    32     97
Sequence      TMHMM2.0      TMhelix    98    120
Sequence      TMHMM2.0      inside   121    132
Sequence      TMHMM2.0      TMhelix   133    155
Sequence      TMHMM2.0      outside   156    200
Sequence      TMHMM2.0      TMhelix   201    223
Sequence      TMHMM2.0      inside   224    267
Sequence      TMHMM2.0      TMhelix   268    290
Sequence      TMHMM2.0      outside   291    304
Sequence      TMHMM2.0      TMhelix   305    327
Sequence      TMHMM2.0      inside   328    343
```

TMHMM posterior probabilities for Sequence



# [plot](#) in postscript, [script](#) for making the plot in gnuplot, [data](#) for plot

# Paralogous Family display

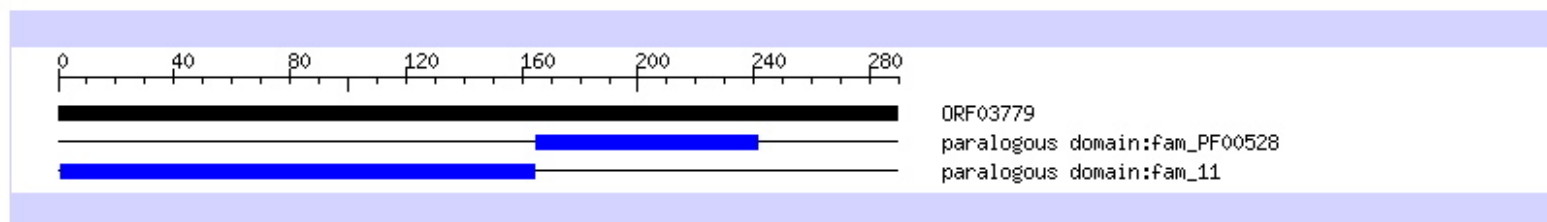
NOTE: this display is for ORF03779

*Shewanella oneidensis* MR-1

Paralogous Families For ORF03779

Logged into [gsp] as [mlgwinm](#)

This page displays information about all paralogous families contained within a specified ORF. The user is presented with three types of information; a graphic display showing the relationship of all paralogous families to the ORF, an ORF list of all ORFs that belong to a paralogous family, and a multiple sequence alignment of all ORFs that belong to a paralogous family.

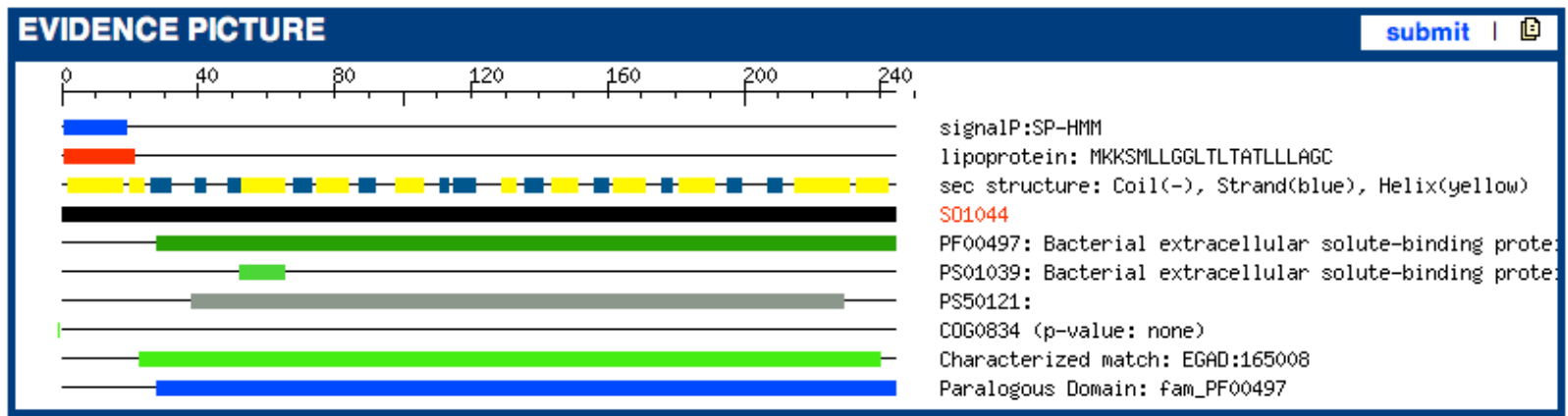


| A | C | Feat_name                          | Common Name                                  | Role_id(s) | <a href="#">11</a> | <a href="#">PF00528</a> | Other Fams |
|---|---|------------------------------------|----------------------------------------------|------------|--------------------|-------------------------|------------|
|   | ● | <a href="#">ORF02518</a><br>283 aa | sulfate ABC transporter, permease protein    | 143        | +                  | +                       |            |
|   | ● | <a href="#">ORF02519</a><br>293 aa | sulfate ABC transporter, permease protein    | 143        | +                  | +                       |            |
|   | ● | <a href="#">ORF02772</a><br>226 aa | molybdenum ABC transporter, permease protein | 143        | +                  | +                       |            |
|   | ● | <a href="#">ORF03459</a><br>245 aa | molybdenum ABC transporter, permease protein | 143        | +                  | +                       |            |
|   | ● | <a href="#">ORF03779</a><br>289 aa | sulfate ABC transporter, permease protein    | 143        | +                  | +                       |            |
|   | ● | <a href="#">ORF03783</a><br>281 aa | sulfate ABC transporter, permease protein    | 143        | +                  | +                       |            |
|   | ● | <a href="#">ORF00271</a><br>343 aa | peptide ABC transporter, permease protein    | 142        |                    | +                       |            |

## Evidence picture from ORF01166

Additional evidence types shown here are signal P, lipoprotein predictions, and PROSITE hits. Signal P and PROSITE information are displayed both in the Evidence Picture and in sections of their own on the Gene Curation Page (next slide). Clicking on the bars in the graphic opens windows with additional information.

Lipoprotein predictions are based on one particular PROSITE motif, so clicking on the red lipoprotein bar will take you to the PROSITE page for the lipoprotein signature (not shown in tutorial).



NOTE: this display is for ORF01166

# Gene Curation Page - PROSITE and Signal P sections on the GCP

NOTE: this display is for ORF01166

Click here to see info on PROSITE motif.

**PROSITE** [submit](#) |

[PS01039](#): Bacterial extracellular solute-binding proteins, family 3 signature.  
Match sequence: **GFDIELAKQIAKDA**

---

| Coords | Precision | Recall | Curation                            |                                      |
|--------|-----------|--------|-------------------------------------|--------------------------------------|
| 52/65  | 0.76      | 0.93   | <input checked="" type="checkbox"/> | <a href="#">[Add To GO Evidence]</a> |

**ATTRIBUTES** [submit](#) |

No Frameshifts Detected.

**SIGNAL\_P** [submit](#) |

SignalP-2.0 Results: [\[Graphical Display\]](#) [\[Raw output for SP-HMM/NN\]](#)

SignalP-2.0 HMM

Prediction  ☐ Curated

Signal peptide probability 0.984

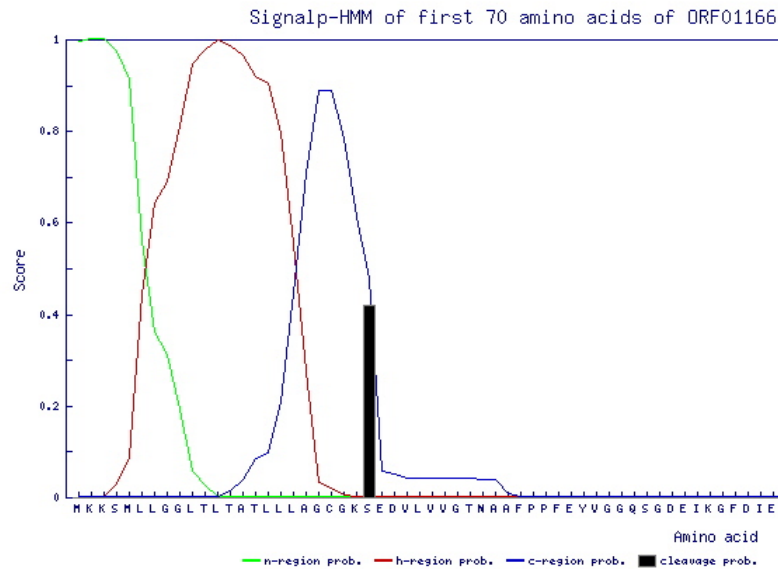
Signal anchor probability

Max cleavage site probability 0.340

Click here to see output in graphical form.

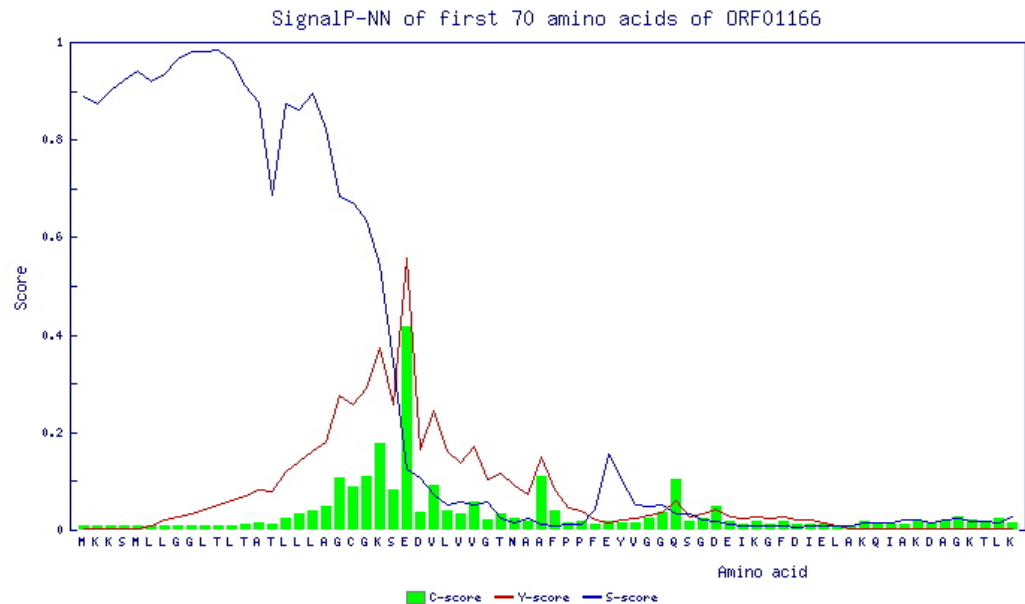
## Signal P Graphical output

NOTE: this display  
is for ORF01166



```
# Using neural networks (NN) and hidden Markov models (HMM) trained on
>ORF01166
Prediction: Signal peptide
Signal peptide probability: 0.993
Signal anchor probability: 0.006
Max cleavage site probability: 0.419 between pos. 24 and 25
```

```
#aa: MKKSMLLGGLTLTATLLLAGCGKS-E
```



```
# Using neural networks (NN) and hidden Markov models (HMM) trained on eukaryotes
# Measure Position Value Cutoff signal peptide?
max. C 25 0.420 0.33 YES
max. Y 25 0.556 0.32 YES
max. S 11 0.981 0.82 YES
mean S 1-24 0.833 0.47 YES
# Most likely cleavage site between pos. 24 and 25:
MKKSMLLGGLTLTATLLLAGCGKS-ED
```

The Korean ExPASy site, [kr.expasy.org](#), is temporarily not available.

Search  for  Go

# PROSITE page at ExPASy

## NiceSite View of PROSITE: [PDOC00798](#) NOTE: this display is for ORF01166 (documentation)

### Bacterial extracellular family 3 signature

#### PROSITE cross-reference(s)

[PS01039: SBP BACTERIAL 3](#)

#### Documentation

Bacterial high affinity transport solutes across the cytoplasmic traffic systems include one of two membrane-associated ATP-binding (PDOC00185) and a high affinity are thought to bind the substrate to transfer it to a complex of the cytoplasm.

In gram-positive bacteria which are surrounded by a single membrane and have therefore no periplasmic region the equivalent proteins are bound to the membrane via an N-terminal lipid anchor. These homolog proteins do not play an integral role in the transport process per se, but probably serve as receptors to trigger or initiate translocation of the solute through the membrane by binding to external sites of the integral membrane proteins of the efflux system.

In addition at least some solute-binding proteins function in the initiation of sensory transduction pathways.

On the basis of sequence similarities, the vast majority of these solute-binding proteins can be grouped into eight families of clusters which generally correlate with the

Family 3 groups together proteins and a periplasmic h

- Histidine-binding protein bacteria. An homologous l
- Lysine/arginine/ornithine coli and related bacter hisJ. Both solute-binding receptor hisP of the bind
- Glutamine-binding protein stearothermophilus.
- Glutamate-binding protein
- Arginine-binding proteins
- Nopaline-binding protein
- Octopine-binding protein
- Major cell-binding factor
- Bacteroides nodosus prote
- Cyclohexadienyl/arogenate periplasmic enzyme which biosynthesis.
- Escherichia coli protein
- Vibrio harveyi protein pa
- Escherichia coli hypothet
- Bacillus subtilis hypothe
- Bacillus subtilis hypothe

The signature pattern is loc

#### Description of pattern(s) and/or profile(s)

|                          |                                                                       |
|--------------------------|-----------------------------------------------------------------------|
| <b>Consensus pattern</b> | G-[FYIL]-[DE]-[LIVMT]-[DE]-[LIVMF]-x(3)-[LIVMA]-[VAGC]-x(2)-[LIVMAGN] |
|--------------------------|-----------------------------------------------------------------------|

|                                                                        |      |
|------------------------------------------------------------------------|------|
| <b>Sequences known to belong to this class detected by the pattern</b> | ALL. |
|------------------------------------------------------------------------|------|

|                                                 |     |
|-------------------------------------------------|-----|
| <b>Other sequence(s) detected in Swiss-Prot</b> | 23. |
|-------------------------------------------------|-----|

#### Last update

November 1997 / Pattern and text revised.

#### References

- [ 1]  
 Tam R., Saier M.H. Jr.  
 Microbiol. Rev. 57:320-346(1993).

#### Copyright

This PROSITE entry is copyright by the Swiss Institute of Bioinformatics (SIB). There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[View entry in original PROSITE document format](#)

[View entry in raw text format \(no links\)](#)

# Gene Curation Page (ORF04813) - Gene Ontology Display

Current GO term assignments are listed in table.

-Click id # to see term in tree.

-Click box for GO term to be deleted.

-Click “add” to add additional evidence rows. (or click delete and add to completely redo evidence)

-Click “edit” to edit evidence.

-“Make ISS”(not seen in this example) can be used when the GO term and evidence assigned by AutoAnnotate are correct, clicking this button marks the old association for deletion and automatically puts in the new info for insertion.

These pull downs have commonly used GO terms. If you choose the unknown terms from any pull-down, the evidence will automatically fill in (since it is always the same.)

Fill in the fields in this section to add or change GO term assignments. All entries must have “ev\_code”, “reference”, and “with”. (more on this in a minute.....)

Link to GO suggestions

Link to GO search tool

**GENE ONTOLOGY**

submit | go sug | search |

| delete                   | go id                                                                   |                              | assigned | date     | evidence                                                         |
|--------------------------|-------------------------------------------------------------------------|------------------------------|----------|----------|------------------------------------------------------------------|
| <input type="checkbox"/> | <a href="#">GO:0004076</a> <a href="#">[add]</a> <a href="#">[edit]</a> | (F) biotin synthase activity | mlgwin   | 07/29/04 | <a href="#">ISS</a> : PMID:12368813 with TIGR_TIGRFAMS:TIGR00433 |
| <input type="checkbox"/> | <a href="#">GO:0009102</a> <a href="#">[add]</a> <a href="#">[edit]</a> | (P) biotin biosynthesis      | mlgwin   | 07/29/04 | <a href="#">ISS</a> : PMID:12368813 with TIGR_TIGRFAMS:TIGR00433 |

function

process

component

| add go id | ev code        | reference                     | with | qualifier |
|-----------|----------------|-------------------------------|------|-----------|
|           | <div>ISS</div> | <div>TIGR_CM:annotation</div> |      |           |
|           | <div>ISS</div> | <div>TIGR_CM:annotation</div> |      |           |
|           | <div>ISS</div> | <div>TIGR_CM:annotation</div> |      |           |
|           | <div>ISS</div> | <div>TIGR_CM:annotation</div> |      |           |
|           | <div>ISS</div> | <div>TIGR_CM:annotation</div> |      |           |

## Overview of steps in GO annotation:

- Review the GO terms assigned to the gene by AutoAnnotate (if any). If they are correct and sufficient use the “Make ISS” button. (not seen here)
- Look for any other needed GO terms in the various suggestion areas on the page: EC#s, HMMs, GO suggestions (see suggestion slide for more info)
- If correct GO terms are unavailable on the Gene Curation Page go to the GO search pages and find the GO term you need. You get there by clicking the search link in the upper right corner of the GO section.
- GO terms must be added in the bottom part of the Gene Ontology section. The GO term id goes in the “add go id” column
- The ec\_code column has a pull-down for choosing the ev\_code you want, the default is “ISS”
- Next is the “reference”, “with”, and “qualifier” columns. Additional slides following this one detail the search for and insertion of GO terms and evidence.
- See the “overview” presentation for more info on GO

GENE ONTOLOGY
submit | go sug | search |

| delete | go id                                                                   |                              | assigned | date     | evidence                                        |
|--------|-------------------------------------------------------------------------|------------------------------|----------|----------|-------------------------------------------------|
|        | <a href="#">GO:0004076</a> <a href="#">[add]</a> <a href="#">[edit]</a> | (F) biotin synthase activity | mlgwin   | 07/29/04 | ISS: PMID:12368813 with TIGR_TIGRFAMS:TIGR00433 |
|        | <a href="#">GO:0009102</a> <a href="#">[add]</a> <a href="#">[edit]</a> | (P) biotin biosynthesis      | mlgwin   | 07/29/04 | ISS: PMID:12368813 with TIGR_TIGRFAMS:TIGR00433 |

function

process

component

| add go id | ev code | reference          | with | qualifier |
|-----------|---------|--------------------|------|-----------|
|           | ISS     | TIGR_CM:annotation |      |           |
|           | ISS     | TIGR_CM:annotation |      |           |
|           | ISS     | TIGR_CM:annotation |      |           |
|           | ISS     | TIGR_CM:annotation |      |           |
|           | ISS     | TIGR_CM:annotation |      |           |

# Gene Curation Page - GO suggestions and Auto-fill-ins

GO term suggestions and auto-fill-in buttons are located in several places on the Gene Curation Page:

- GO terms assigned to [HMMs](#) are listed under HMM hits (if any have been assigned - see the HMM slide for how these look). These are often excellent sources for GO terms. Clicking the “Add” button next to a GO term under an HMM adds both the term id and the evidence to the appropriate fields in the GO entry section. Clicking the “Add to GO evidence” button adds just the HMM accession into the “with” field in the GO entry section.

- GO terms corresponding to [EC numbers](#) are listed next to the EC box (for enzymes). Clicking the “add” button will put the GO term id into the “add go id” fields in the GO entry section.

- GO terms assigned manually to [other bacterial genomes](#) (V. cholerae, B. anthracis - both a Gram + and Gram - representative), [InterPro hits](#), [Genome Properties](#) are listed both at the bottom of the page and in a pop-up window accessed by the link in the upper right corner of the GO section. Clicking on “add” in this section puts the GO id into the “add go id” fields in the GO entry section.

- “Add to GO evidence” buttons are also available for [Prosites](#) hits, this populates the “with” field with the Prosite accession. Available when a protein has matches to Prosite.

- “Add to GO evidence” is also available for the [characterized match accession](#), this will put the accession of the characterized matching protein into the “with” field entry box.

- “Add to GO evidence” is also available for [Genome Properties](#), clicking on the “GO” link under the “add to GO evidence” column in the Genome Properties section will enter the GenProp accession in the “with” field.

See next page for screen shots.

GO terms and evidence  
Auto Fill-ins

GENE ONTOLOGY

submit | go sug | search |

| delete                   | go id                                               |                              | assigned | date     | evidence                                        |
|--------------------------|-----------------------------------------------------|------------------------------|----------|----------|-------------------------------------------------|
| <input type="checkbox"/> | GO:0004076 <a href="#">add</a> <a href="#">edit</a> | (F) biotin synthase activity | mlgwin   | 07/29/04 | ISS: PMID:12368813 with TIGR_TIGRFAM5:TIGR00433 |
| <input type="checkbox"/> | GO:0009102 <a href="#">add</a> <a href="#">edit</a> | (P) biotin biosynthesis      | mlgwin   | 07/29/04 | ISS: PMID:12368813 with TIGR_TIGRFAM5:TIGR00433 |

function

process

component

add go id

evcode

reference

with

qualifier

|                      |     |                    |                      |                      |                      |
|----------------------|-----|--------------------|----------------------|----------------------|----------------------|
| <input type="text"/> | ISS | TIGR_CM:annotation | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| <input type="text"/> | ISS | TIGR_CM:annotation | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| <input type="text"/> | ISS | TIGR_CM:annotation | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| <input type="text"/> | ISS | TIGR_CM:annotation | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| <input type="text"/> | ISS | TIGR_CM:annotation | <input type="text"/> | <input type="text"/> | <input type="text"/> |

EC GO suggestions:

GO:0004076  biotin synthase activity (function)

GO SUGGESTIONS:

InterPro GO Suggestions:

IPR02684: [Add To GO Evidence](#)

- GO:0004076  biotin synthase activity (function)
- GO:0009102  biotin biosynthesis (process)

Vibrio Cholera GO Suggestions:

ORF01454 biotin synthase (pvalue = 6.3e-121)

- GO:0004076  biotin synthase activity (function)
- GO:0009102  biotin biosynthesis (process)

Bacillus anthracis GO Suggestions:

ORF03371 biotin synthase (pvalue = 8.3e-48)

- GO:0004076  biotin synthase activity (function)
- GO:0009102  biotin biosynthesis (process)

Genome Property GO Suggestions:

GO:0009102  biotin biosynthesis (process) | biotin biosynthesis (YES)

CHARACTERIZED MATCH

submit | history |

(aln) SP:P12996 coords: 7 / 350 score: 42 Pvalue: 7.2e-120 per\_id: 66.0% per\_sim: 79.7% [Add To GO Evidence](#)

Delete accession:  Add accession:  [Add To GO Evidence](#)

HMM

submit | all hmms |

TIGR00433: biotin synthase

Isology: **equivalog**

Total score: 564.1 Trusted cutoff: 300.00 Gathering cutoff: 300.00 Noise cutoff: 50.00 Total expect: 1.5e-166

Trusted cutoff 2: 300.00 Gathering cutoff 2: 300.00 Noise cutoff 2: 50.00

View Alignment

align page

| Coords | HMM Coords | Score | Expect   | Curation                            |
|--------|------------|-------|----------|-------------------------------------|
| 18-313 | 1-350/350  | 564.1 | 1.5e-166 | <input checked="" type="checkbox"/> |

GO:0004076  biotin synthase activity (F)

GO:0009102  biotin biosynthesis (P)

Genome Properties

state property name

YES biotin biosynthesis

Run Rules.spl...

add GO evidence

[GO]

## Manatee's GO ontology and annotation search tool:

In many cases the GCP will not have a suggested GO term that meets an annotators needs. In that situation the annotator will turn to Manatee's built in GO ontology and annotation browser.

There are several available functions in the search tool:

- GO term id search of ontologies - this returns a tree view of the search term in the ontology
- GO term name keyword search of ontologies - this returns a table of terms where the name, or a synonym of the name contains the keyword or where a word contains the keyword in question.
- protein name keyword search of annotations - this is a search of the annotations and returns proteins whose name match the keyword and the GO terms that were assigned to those proteins
- GO id search of annotations - search GO annotations with a GO id and see a list of proteins that have been annotated to that GO term.
- GO correlations in annotations - often a particular function term will often be assigned with a particular process term (for example: “biotin synthase” will almost always be assigned in conjunction with “biotin biosynthesis”) - when one needs help finding a process or function one can search for these relationships with the correlations tool.
- EC search - uses the ec2go mapping file provided on the GO web site to look up GO terms that correspond to EC numbers
- GO BLAST - search a protein sequence against a database of proteins that have been annotated to GO, then link to the GO terms that were assigned to them. This is the only GO search tool not accessible on the GO search page - this one is found in the “Select Function” pull-down menu at the top of the Gene Curation Page.

See next page for screen shots.

# Links from the Gene Curation page - GO Search Tool

Click on the “search” link in the title bar of the GO data section

Input a GO term here. This results in a GO term information page.

Input an EC number and get the corresponding GO term

| search GO from GO id:                                                                                                                            | search GO from EC number:                                                                                                                                                                           | search GO from keywords                                                                                                                                                                                                                                                                          | search GO associations with gene name                                                              |
|--------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| GO id: <input type="text"/> <input type="button" value="submit"/> <input type="button" value="reset"/>                                           | EC #: <input type="text"/> <input type="button" value="submit"/> <input type="button" value="reset"/>                                                                                               | <input type="checkbox"/> Exact Match<br><input type="button" value="submit"/> <input type="button" value="reset"/><br><input checked="" type="checkbox"/> Molecular Function<br><input checked="" type="checkbox"/> Biological Process<br><input checked="" type="checkbox"/> Cellular Component | <input type="text"/><br><input type="button" value="submit"/> <input type="button" value="Reset"/> |
| search go_gene_association from GO id:<br>GO id: <input type="text"/> <input type="button" value="submit"/> <input type="button" value="reset"/> | search for GO correlations:<br>GO id: <input type="text"/> <input type="button" value="submit"/> <input type="button" value="reset"/><br><input type="checkbox"/> Search TIGR prokaryotic data only |                                                                                                                                                                                                                                                                                                  |                                                                                                    |

Search for GO terms that most frequently are assigned to a protein along with the input GO term. Check the box to restrict the search to TIGR prokaryotic data only.

Input a search string here using the checkboxes to select some or all of the ontologies to search in. If you want to restrict your results to only terms which share all of your input text, click “Exact match”. This searches the name of the GO term.

Input a GO id and see genes from other databases that have been annotated with that GO term

Input text and see GO terms assigned to genes from other databases whose common names contain the input text.

# GO Term information page and tree view.

This page is reached by clicking on GO id links or using GO id search.

Name: **biotin synthase activity**

Type: **molecular\_function**

Definition: **Catalysis of the reaction\: dethiobiotin + sulfur = biotin.**

Comment: **NONE**

Synonym: **NONE**

Secondary ID: **NONE**

EC Number: **2.8.1.6**

## Absolute Path in GO Tree: 1 instance detected

```
+Ontology (TI:0000001)[R]
+Gene_Ontology (GO:0003673)[P]
+molecular_function (GO:0003674)[P]
+catalytic activity (GO:0003824)[I]
+transferase activity (GO:0016740)[I]
+transferase activity, transferring sulfur-containing groups (GO:0016782)[I]
+sulfurtransferase activity (GO:0016783)[I]
  biotin synthase activity (GO:0004076)[I]
```

Numbers next to the terms in the tree indicate the number of genes from this organism that are annotated to that term or a child of that term - clicking on the number gives you a table of those genes and relevant info. (missing in this screen shot)

If you reach this page by clicking on a GO term on a GCP, clicking the “add” button in the tree will place that GO term in the “add” field on the GCP.

## View Mode: Regular

```
+Ontology (TI:0000001)[R] [add]
+Gene_Ontology (GO:0003673)[P] [add]
+molecular_function (GO:0003674)[P] [add]
+catalytic activity (GO:0003824)[I] [add]
+transferase activity (GO:0016740)[I] [add]
+transferase activity, transferring sulfur-containing groups (GO:0016782)[I] [add]
+sulfurtransferase activity (GO:0016783)[I] [add]
  biotin synthase activity (GO:0004076)[I] [add]
  cysteine desulfurase activity (GO:0031071)[I] [add]
  3-mercaptopyruvate sulfurtransferase activity (GO:0016784)[I] [add]
  tRNA sulfurtransferase activity (GO:0016227)[I] [add]
  thiosulfate-thiol sulfurtransferase activity (GO:0050337)[I] [add]
  thiosulfate-dithiol sulfurtransferase activity (GO:0047362)[I] [add]
  thiosulfate sulfurtransferase activity (GO:0004792)[I] [add]
+transferase activity, transferring alkylthio groups (GO:0050497)[I] [add]
+CoA-transferase activity (GO:0008410)[I] [add]
+sulfotransferase activity (GO:0008146)[I] [add]
+transferase activity, transferring phosphorus-containing groups (GO:0016772)[I] [add]
  pyruvyltransferase activity (GO:0046919)[I] [add]
  CDP-alcohol phosphotransferase activity (GO:0008414)[I] [add]
  trichothecene 3-O-acetyltransferase activity (GO:0045462)[I] [add]
+transferase activity, transferring alkyl or aryl (other than methyl) groups (GO:0016757)[I] [add]
+transferase activity, transferring glycosyl groups (GO:0016757)[I] [add]
+2'-phosphotransferase activity (GO:0008665)[I] [add]
+glucanoyltransferase activity (GO:0042123)[I] [add]
+transferase activity, transferring nitrogenous groups (GO:0016769)[I] [add]
+transferase activity, transferring selenium-containing groups (GO:0016785)[I] [add]
  mannosylphosphate transferase activity (GO:0000031)[I] [add]
+transferase activity, transferring one-carbon groups (GO:0016741)[I] [add]
  cobinamide phosphate guanylyltransferase activity (GO:0008820)[I] [add]
+transferase activity, transferring aldehyde or ketonic groups (GO:0016744)[I] [add]
  lipoyltransferase activity (GO:0017118)[I] [add]
+transferase activity, transferring acyl groups (GO:0016746)[I] [add]
  S-adenosylmethionine:tRNA ribosyltransferase-isomerase activity (GO:0051075)[I] [add]
  lauroyl transferase activity (GO:0008913)[I] [add]
```

# Search results for GO term keyword: “biotin”

The first part of the table shows results from the GO term names.

The second part of the table shows results from GO term synonyms.

Note that areas of the text which matched the keyword are highlighted in red by Manatee.

Terms which are “obsolete” or “secondary” to another term will have that indicated in column one.

Click any GO term id number for a view of the term in the GO tree.

| go_term hits           |                    |                                                                      |
|------------------------|--------------------|----------------------------------------------------------------------|
| GO id                  | type               | name                                                                 |
| GO:0009374             | molecular_function | biotin binding                                                       |
| GO:0009102             | biological_process | biotin biosynthesis                                                  |
| GO:0042966             | biological_process | biotin carboxyl carrier protein biosynthesis                         |
| GO:0004075             | molecular_function | biotin carboxylase activity                                          |
| GO:0009343             | cellular_component | biotin carboxylase complex                                           |
| GO:0042367             | biological_process | biotin catabolism                                                    |
| GO:0006768             | biological_process | biotin metabolism                                                    |
| GO:0004076             | molecular_function | biotin synthase activity                                             |
| GO:0015878             | biological_process | biotin transport                                                     |
| GO:0015225             | molecular_function | biotin transporter activity                                          |
| GO:0047707             | molecular_function | biotin-CoA ligase activity                                           |
| GO:0004077             | molecular_function | biotin-[acetyl-CoA-carboxylase] ligase activity                      |
| GO:0004078             | molecular_function | biotin-[methylcrotonoyl-CoA-carboxylase] ligase activity             |
| GO:0004079             | molecular_function | biotin-[methylmalonyl-CoA-carboxyltransferase] ligase activity       |
| GO:0004080             | molecular_function | biotin-[propionyl-CoA-carboxylase (ATP-hydrolyzing)] ligase activity |
| GO:000106<br>secondary | molecular_function | biotin-apoprotein ligase activity                                    |
| GO:0018271             | molecular_function | biotin-protein ligase activity                                       |
| GO:0047708             | molecular_function | biotinidase activity                                                 |
| GO:0019351             | biological_process | dethiobiotin biosynthesis                                            |
| GO:0046450             | biological_process | dethiobiotin metabolism                                              |
| GO:0004141             | molecular_function | dethiobiotin synthase activity                                       |
| GO:0018054             | biological_process | peptidyl-lysine biotinylation                                        |
| GO:0009305             | biological_process | protein amino acid biotinylation                                     |

| go_synonym hits |                    |                                                                |                                   |
|-----------------|--------------------|----------------------------------------------------------------|-----------------------------------|
| GO id           | type               | name                                                           | synonym                           |
| GO:0004079      | molecular_function | biotin-[methylmalonyl-CoA-carboxyltransferase] ligase activity | biotin-1                          |
| GO:0018271      | molecular_function | biotin-protein ligase activity                                 | biotin-apoprotein ligase activity |
| GO:0019351      | biological_process | dethiobiotin biosynthesis                                      | dethiobiotin biosynthesis         |
| GO:0046450      | biological_process | dethiobiotin metabolism                                        | dethiobiotin metabolism           |

# GO correlations search

## Search results from query with GO:0004076 “biotin synthase activity”

Searches data set stored in our database of all associations to genes available on GO web site.

First table shows percentages of occurrences of the query term with other terms.

Second table shows details of all instances of query term assigned to a gene in the data set.

### GO Correlations: GO:0004076 | biotin synthase activity

Logged into [gsp] as [mlgwinn](#)



| correlation percentage | GO id                      | GO name                    | GO type |
|------------------------|----------------------------|----------------------------|---------|
| 88.89 %                | <a href="#">GO:0009102</a> | biotin biosynthesis        | P       |
| 11.11 %                | <a href="#">GO:0008372</a> | cellular_component unknown | C       |

| gene id   | gene db   | gene name                      | GO id      | GO type | GO name                    |
|-----------|-----------|--------------------------------|------------|---------|----------------------------|
| S0003518  | SGD       | biotin synthase                | GO:0008372 | C       | cellular_component unknown |
| S0003518  | SGD       | biotin synthase                | GO:0009102 | P       | biotin biosynthesis        |
| BA4336    | gba_TIGR  | biotin synthetase              | GO:0009102 | P       | biotin biosynthesis        |
| SO3925    | gsp_TIGR  | biotin synthase family protein | GO:0009102 | P       | biotin biosynthesis        |
| SO2740    | gsp_TIGR  | biotin synthase                | GO:0009102 | P       | biotin biosynthesis        |
| CBU1007   | gcb_TIGR  | biotin synthase                | GO:0009102 | P       | biotin biosynthesis        |
| VC1112    | gvc_TIGR  | biotin synthase                | GO:0009102 | P       | biotin biosynthesis        |
| GSORF1608 | ggs_TIGR  | biotin synthetase              | GO:0009102 | P       | biotin biosynthesis        |
| At2g43360 | TIGR_Ath1 | biotin synthase (BioB) (BIO2)  | GO:0009102 | P       | biotin biosynthesis        |

# Output from GO search for protein common name keyword: biotin synthase

| gene id     | role id | gene symbol | EC #    | GO id (GO type) | ev code | GO term                                  | gene name                      |
|-------------|---------|-------------|---------|-----------------|---------|------------------------------------------|--------------------------------|
| SO3925      | 157     |             |         | GO:0004076 (F)  | ISS     | biotin synthase activity                 | biotin synthase family protein |
| SO3925      | 157     |             |         | GO:0009102 (P)  | ISS     | biotin biosynthesis                      | biotin synthase family protein |
| SO2737      | 77      | bioD        | 6.3.3.3 | GO:0004141 (F)  | ISS     | dethiobiotin synthase activity           | dethiobiotin synthase          |
| SO2737      | 77      | bioD        | 6.3.3.3 | GO:0009102 (P)  | ISS     | biotin biosynthesis                      | dethiobiotin synthase          |
| SO2740      | 77      | bioB        | 2.8.1.6 | GO:0004076 (F)  | ISS     | biotin synthase activity                 | biotin synthase                |
| SO2740      | 77      | bioB        | 2.8.1.6 | GO:0009102 (P)  | ISS     | biotin biosynthesis                      | biotin synthase                |
| CBU1007     | 77      | bioB        | 2.8.1.6 | GO:0004076 (F)  | ISS     | biotin synthase activity                 | biotin synthase                |
| CBU1007     | 77      | bioB        | 2.8.1.6 | GO:0009102 (P)  | ISS     | biotin biosynthesis                      | biotin synthase                |
| VC1112      | 77      | bioB        | 2.8.1.6 | GO:0004076 (F)  | ISS     | biotin synthase activity                 | biotin synthase                |
| VC1112      | 77      | bioB        | 2.8.1.6 | GO:0009102 (P)  | ISS     | biotin biosynthesis                      | biotin synthase                |
| GSU1583     | 77      | bioD        | 6.3.3.3 | GO:0004141 (F)  | ISS     | dethiobiotin synthase activity           | dethiobiotin synthase          |
| GSU1583     | 77      | bioD        | 6.3.3.3 | GO:0009102 (P)  | ISS     | biotin biosynthesis                      | dethiobiotin synthase          |
| P32451      |         |             |         | GO:0005739 (C)  | IDA     | mitochondrion                            | Biotin synthase                |
| Q84QK2      |         |             |         | GO:0004076 (F)  | IEA     | biotin synthase activity                 | Putative biotin synthase       |
| Q84QK2      |         |             |         | GO:0005739 (C)  | ISS     | mitochondrion                            | Putative biotin synthase       |
| Q84QK2      |         |             |         | GO:0009102 (P)  | IEA     | biotin biosynthesis                      | Putative biotin synthase       |
| At2g43360   |         |             |         | GO:0004076 (F)  | IGI     | biotin synthase activity                 | biotin synthase (BioB) (BIO2)  |
| At2g43360   |         |             |         | GO:0009102 (P)  | TAS     | biotin biosynthesis                      | biotin synthase (BioB) (BIO2)  |
| SPCC1235.02 |         |             |         | GO:0004076 (F)  | IEA     | biotin synthase activity                 | biotin synthase activity       |
| SPCC1235.02 |         |             |         | GO:0006731 (P)  | IEA     | coenzyme and prosthetic group metabolism | biotin synthase activity       |
| SPCC1235.02 |         |             |         | GO:0006790 (P)  | IEA     | sulfur metabolism                        | biotin synthase activity       |
| S0003518    |         |             |         | GO:0004076 (F)  | TAS     | biotin synthase activity                 | biotin synthase                |
| S0003518    |         |             |         | GO:0005739 (C)  | IDA     | mitochondrion                            | biotin synthase                |
| S0003518    |         |             |         | GO:0009102 (P)  | TAS     | biotin biosynthesis                      | biotin synthase                |

# Adding GO Evidence

**Step 1. Pick an evidence code.** Most genes in bacterial genome sequencing projects will get an ev\_code of “ISS”. This stands for “Inferred from sequence similarity.” If a gene from the sequenced organism has had experimental characterization, then chose an appropriate experimental ev\_code. All “unknown” GO terms get “ND” as ev\_code. To see all ev\_codes, click the “ev\_code” link.

**Step 2. Fill in “reference” information.** For ISS terms prior to publication use “TIGR\_CM:annotation”, after publication use the

PMID of the genome paper. For “unknown” terms use “GO\_REF:nd”. For terms with experimental evidence codes use the PMID of the paper describing the characterization.

**Step 3. Fill in the “with” field.** For all ISS entries you must fill in the accession of the HMM, characterized match protein, or Genome Property that led to the annotation.

**Auto Fill-ins:** Both the GO ids and associated evidence can be filled in automatically by clicking the “Add” buttons next to GO suggestions and the “Add to GO evidence” buttons. All info for the “unknown” terms is filled in automatically by choosing the “unknown” terms in the pull-down menus. All information for GO terms assigned to HMMs is filled in with the “Add” buttons next to GO terms under HMMs.

**Qualifier** should be set to “contributes\_to” when annotating the function of a complex to the proteins that make up the complex. (see the overview for more informatin on all of these fields)

GENE ONTOLOGY
submit | go sug | search |

| delete                   | go id                                                               |                              | assigned by | assign date | evidence                                                                                                                       |
|--------------------------|---------------------------------------------------------------------|------------------------------|-------------|-------------|--------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> | <a href="#">GO:0004076</a> <a href="#">add</a> <a href="#">edit</a> | (F) biotin synthase activity | mlgwin      | 03/29/04    | <a href="#">ISS</a> : PMID:12368813 with Swiss-Prot:P12996<br><a href="#">ISS</a> : PMID:12368813 with TIGR_TIGRFAMS:TIGR00433 |
| <input type="checkbox"/> | <a href="#">GO:0009102</a> <a href="#">add</a> <a href="#">edit</a> | (P) biotin biosynthesis      | mbeanan     | 11/15/01    | <a href="#">ISS</a> : PMID:12368813 with Swiss-Prot:P12996<br><a href="#">ISS</a> : PMID:12368813 with TIGR_TIGRFAMS:TIGR00433 |

function

process

component

| add go id | ev code | reference          | with | qualifier |
|-----------|---------|--------------------|------|-----------|
|           | ISS     | TIGR_CM:annotation |      |           |
|           | ISS     | TIGR_CM:annotation |      |           |
|           | ISS     | TIGR_CM:annotation |      |           |
|           | ISS     | TIGR_CM:annotation |      |           |
|           | ISS     | TIGR_CM:annotation |      |           |

# Gene Curation Page - TIGR roles

Click here to view/edit role notes

Click here to enter this role into the "Delete" box

Click on the name of the main role or sub role to take you to a page with the gene list for that main/sub role.

| TIGR ROLES         |                     |                                                                            |                        | <a href="#">submit</a> | <a href="#">role help</a> | <a href="#">history</a> |
|--------------------|---------------------|----------------------------------------------------------------------------|------------------------|------------------------|---------------------------|-------------------------|
| role_id            | delete              | main role                                                                  | sub role               |                        |                           |                         |
| <a href="#">77</a> | <a href="#">del</a> | <a href="#">Biosynthesis of cofactors, prosthetic groups, and carriers</a> | <a href="#">Biotin</a> |                        |                           |                         |

Add role\_ids (separate with spaces):

Delete role\_ids (click on ids above):

Add or delete role ids with these boxes.

Click here for a list of TIGR roles.

## Gene Curation Page - How to get the data into the database: The “Submit” buttons

The screenshot shows a web interface titled "SUBMIT DATA" in a dark blue header. Below the header, there is a light blue box containing the text "Start confidence not calculated." and two checked checkboxes: "Start Site Curated:rdeboy" and "Completed: rdeboy". To the right of these checkboxes are two buttons: "Submit" and "Reset".

Click this button when you have completed annotation for this gene. With this toggle we know that this gene is finished.

Clicking this button indicates that you have reviewed the start site and either found it to be fine or edited it to the correct (or at least what we hope is correct) position.

Click here to submit your entries to the database. You can also do this by clicking on any of the “submit” buttons in the upper right of each section on the page. Clicking “submit” anywhere on the page submits data from all fields (not just the section from which you clicked the button.)

This button resets the page to the state it was when originally opened.

## Gene Curation Page - The pull down menus

If you click on the select pull down menus you will get a selection of options. Each of these when selected will generate a new page with the desired information. (Later slides show examples of some of these.)

Shewanella oneidensis MR-1

Gene Curation Page

[Home](#) | [Logged into \[gsp\] as mlgwinn](#)

**ORF04813 (SO2740)**  
▶ [View BER Searches](#)  
asmb\_id: 7974  
▶ [Reload Page](#)

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

database:   
feat\_name / locus:

Select Display  
Select Display  
Genome Region  
View Sequences  
3rd Position GC Skew  
Signal Peptide Prediction  
Transmembrane Helix Prediction  
Secondary Structure Prediction

Select Function  
Select Function  
Edit Start Sites  
Frameshift Report  
Translation Exception  
Make This ORF Hypothetical  
Blast Gene Against GO

GEI

gene

bioti

gene\_sym:  
bioB

EC number(s):  
 

comment:

▶ [nt\\_comment](#)

| [history](#) | 

EC GO suggestions:  
▶ [GO:0004076](#)  biotin synthase activity (F)

pub\_comment:

▶ [auto\\_comment](#)

## Links from the Gene Curation Page - View sequence

This page shows the length (nucleotide and protein), coordinates, MW, and pl of the protein.

Also, in fasta format are the nucleotide and protein sequences.

**length:** 1050 nucleotides  
**protein length:** 350 amino acids  
**MW:** 38790.13  
**pl:** 4.9477

### Genomic sequence

```
>ORF04813
ATGTCGCAAGTTGCAAGTTGTCATGATTGGAAGCGGGAAGAAATCGAAGCCTTATTTGCG
CTGCCGATGAATGACTTATTATTTAAAGCCCACAGTATCCACCGTGAAGAGTACGATCCT
AACGAAAGTGCAGATCAGCCGCTTATTGTCGATCAAAACTGGGGCTTGTCCTGAGGATTGT
AAATATTTGTCGCGAGAGTGCAGCTTACGACACTGGCCCTTGAAAAAGAGCGTCTCTTAGCG
ATGGAAACCGTGCTCACCAGAGCGCGTAGCGCGAAAGCGGCGGGCGCTTCGCGTTTCTGT
ATGGGCGCGCGCTTGGCGTAACCCGAAAGATAAAGATATGCCATACCTCAAGCAAATGGTG
CAAGAGGTGAAAGCCCTCGGCATGGAAACCTGTATGACCTTAGGGATGTTAAGTGCCGAG
CAAGCCAATGAGTTGGCCGAGCAGGCCCTTGACTATTACAACCACAATTTAGATACCTCG
CCTGAATACTACGGCGATGTGATCACCACCCGTACCTATCAAAACCGCTTAGATACCTTA
AGCCATGTGCGCGCATCGGGCATGAAAGTTTGCTCTGGCGGCATTGTGCGCATGGGCGAG
AAGGCTACTGACAGAGCCGGTTTATTACAACAACTGGCTAATTTACCCAGCATCCGGAT
TCTGTGCCGATCAATATGTTAGTCAAAGTAGCGGGTACCCCTTTGAAAAACTTGATGAT
TTAGATCCACTCGAGTTTGTCCGAAACCATCGCCGTGGCGCGTATTTTAATGCCACTGTGCG
CGGGTGCGTTTATCCGCGAGGCCGTGAAAAATATGAGCGATGAGTGCAGGCCATGTGTTTC
TTTGCGGGCGCGAACTCGATTTTTTACGGCTGTAAAGTTACTGACCACGCCCCAACCCGAA
GAAAGTGATGATATGGGGTTGTTCCCGTCCGCTGGGTTTACGCCCTGAGCAGGGCGCAGCC
GCCTCTATTGATGATGAGCAAGCGGTATTAGCTAAAGCTGCGGCTTATCAAGATAAAGCT
TCAGCTCAGTTTTATGATGCGGCGGCACTA
```

### CDS

```
>ORF04813
ATGTCGCAAGTTGCAAGTTGTCATGATTGGAAGCGGGAAGAAATCGAAGCCTTATTTGCG
CTGCCGATGAATGACTTATTATTTAAAGCCCACAGTATCCACCGTGAAGAGTACGATCCT
AACGAAAGTGCAGATCAGCCGCTTATTGTCGATCAAAACTGGGGCTTGTCCTGAGGATTGT
AAATATTTGTCGCGAGAGTGCAGCTTACGACACTGGCCCTTGAAAAAGAGCGTCTCTTAGCG
ATGGAAACCGTGCTCACCAGAGCGCGTAGCGCGAAAGCGGCGGGCGCTTCGCGTTTCTGT
ATGGGCGCGCGCTTGGCGTAACCCGAAAGATAAAGATATGCCATACCTCAAGCAAATGGTG
CAAGAGGTGAAAGCCCTCGGCATGGAAACCTGTATGACCTTAGGGATGTTAAGTGCCGAG
CAAGCCAATGAGTTGGCCGAGCAGGCCCTTGACTATTACAACCACAATTTAGATACCTCG
CCTGAATACTACGGCGATGTGATCACCACCCGTACCTATCAAAACCGCTTAGATACCTTA
AGCCATGTGCGCGCATCGGGCATGAAAGTTTGCTCTGGCGGCATTGTGCGCATGGGCGAG
AAGGCTACTGACAGAGCCGGTTTATTACAACAACTGGCTAATTTACCCAGCATCCGGAT
TCTGTGCCGATCAATATGTTAGTCAAAGTAGCGGGTACCCCTTTGAAAAACTTGATGAT
TTAGATCCACTCGAGTTTGTCCGAAACCATCGCCGTGGCGCGTATTTTAATGCCACTGTGCG
CGGGTGCGTTTATCCGCGAGGCCGTGAAAAATATGAGCGATGAGTGCAGGCCATGTGTTTC
TTTGCGGGCGCGAACTCGATTTTTTACGGCTGTAAAGTTACTGACCACGCCCCAACCCGAA
GAAAGTGATGATATGGGGTTGTTCCCGTCCGCTGGGTTTACGCCCTGAGCAGGGCGCAGCC
GCCTCTATTGATGATGAGCAAGCGGTATTAGCTAAAGCTGCGGCTTATCAAGATAAAGCT
TCAGCTCAGTTTTATGATGCGGCGGCACTA
```

### Protein

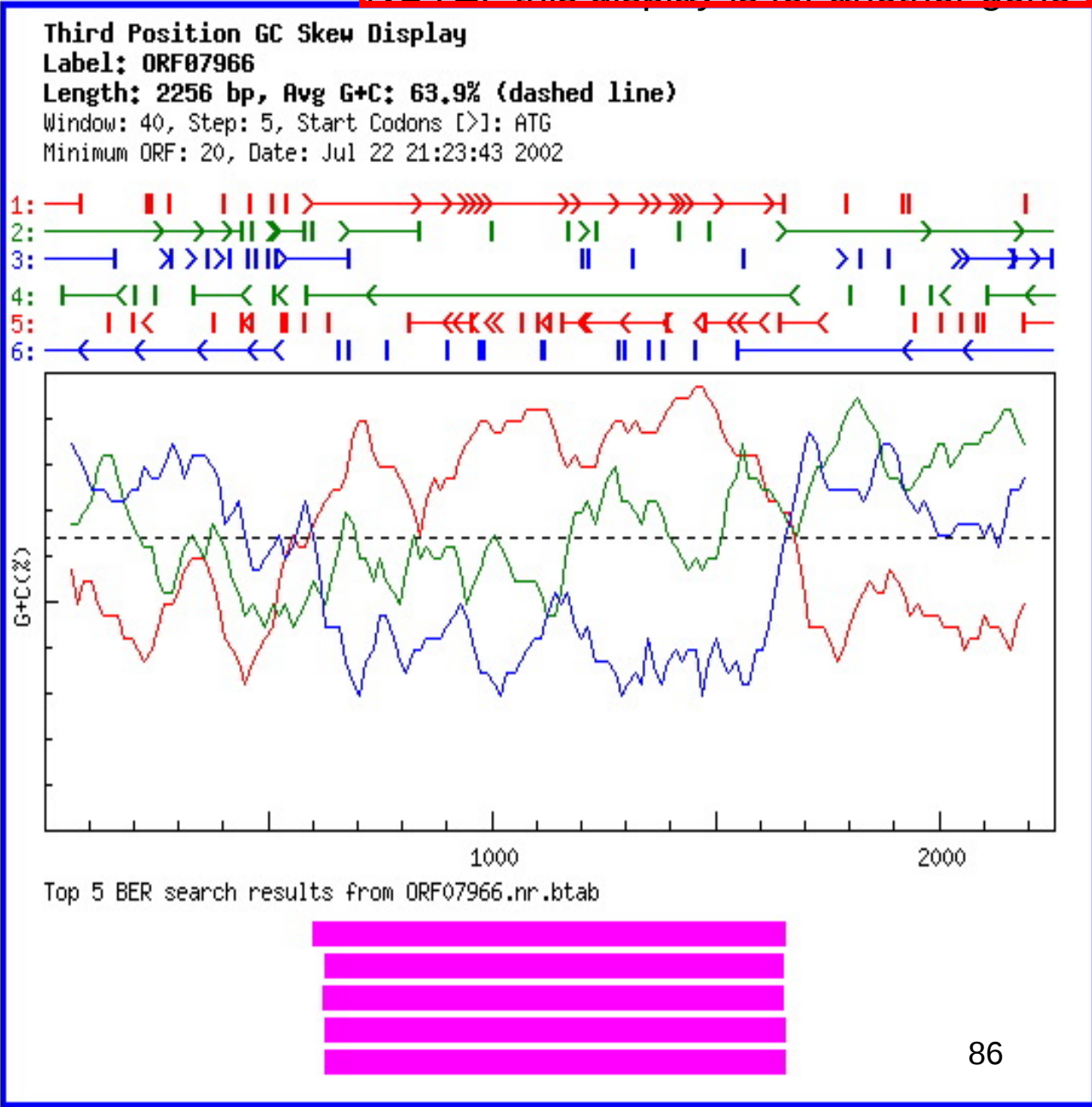
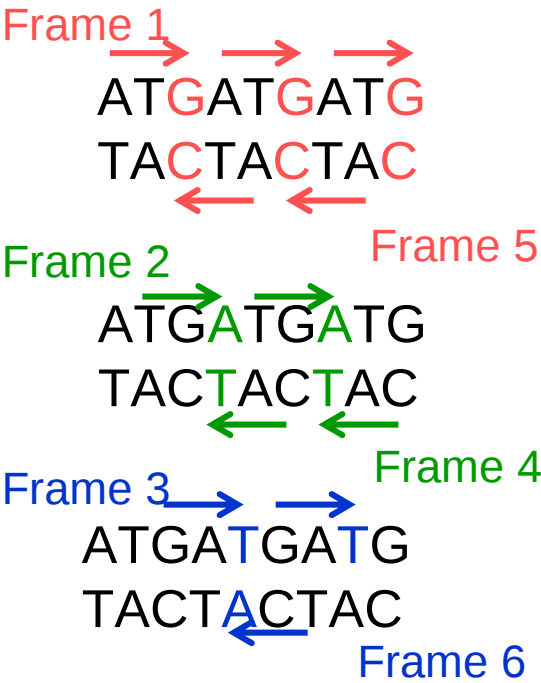
```
>ORF04813
MSQLQVRHDKREEIEALFALPHNDLLFKAHSIHREEYDPNEVQISRLLSIKTGACPEDC
KYCPQSARYIDTGLEKERLLANETVLTARSAAAGASRFCHGAAWRNPKDKDHPYLKQHV
QEVKALGHETCHTLGHL SAEQANELAEAGLDYYNHNDLT SPEYVGDVITTRTYQNRDLTL
SHVRASTHETVSGGIIVGHEKATDRAGLLQQLANLPHQPDVSVFINMLVKVAGTFFEKLLD
LDPLEFVRTIAVARILMPLSRVRLSAGRENMSDELQAMCFAGANSIFVAGCKLLTTPNPE
ESDDMGLFRRLGLRPEQGAASIDDEQAVLAKAAAYQDKASAQFYDAAL
```

# Links from the Gene Curation Page - Third position GC skew

NOTE: this display is for another gene

In organisms whose DNA has a high GC content it can sometimes be helpful to look at third position GC skew to help resolve overlaps.

Due to the nature of the genetic code, the third position is the least constrained of a codon and therefore will be able to reflect the higher GC content of the overall genome. Therefore one should see a markedly higher GC content in the third position of the correct frame.



# ORF Management in Manatee: Genome Viewer

[Refresh XML](#)
[Search](#)

Asmb1 Id: 7974
 Database: gsp

database: 
 asmb1\_id: 
 submit
 reset

| feat_name            | end5                 | end3                 | role_id              | ec_num               | gene_sym             | complete             | com_name             |
|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> |

Six Frame Options (on six frame clicks)

[View Sequence](#)
[Blast](#)
[Insert Gene](#)

Gene Options (on feat\_name clicks)

[View Sequence](#)
[Annotate ORF](#)
[Edit Start](#)
[Blast ORF](#)
[Merge Genes](#)
[Delete Gene](#)

0.0kb 1.0kb 2.0kb 3.0kb 4.0kb 5.0kb 6.0kb 7.0kb 8.0kb

Clicking on the **“Genome Viewer”** option on the “Welcome to Manatee” Page, selecting “GV” next to a gene id in a gene list, or selecting “Genome Region” in the “Select Display” pull-down on the GCP will take you to our Genome Viewer tool. Here you can view the genes from the whole genome in relation to each other, edit their starts, merge them, insert new genes, and delete genes. Mousing over the genes fills in the information boxes near the top of the display with coordinates, com name, etc.

Search

Coordinate:

Search

Lower Coordinate:

Upper Coordinate:

Search

feat\_name:

Search

To get to a specific region of the genome, enter coordinates or feat\_name in the search section at the bottom of every Genome Viewer page. One can also use the pull-down menu on the Gene Curation Page or the “GV” link on a gene list page.

# ORF Management in Manatee - 6-frame analysis

[Refresh XML](#) [Search](#) AsmbI Id: 7974 Database: gsp database: asmbI\_id: submit reset

| feat_name | end5    | end3    | role_id | ec_num | gene_sym | complete | com_name |
|-----------|---------|---------|---------|--------|----------|----------|----------|
|           | 2861863 | 2858294 |         |        |          |          |          |

Six Frame Options (on six frame clicks)  
View Sequence ☒ Blast ☐ Insert Gene ☐

Gene Options (on feat\_name clicks)  
View Sequence ☐ Annotate ORF ☒ Edit Start ☐ Blast ORF ☐ Merge Genes ☐ Delete Gene ☐

To analyze regions in the 6-frame translation (options boxed in pink), click on the button for the activity you wish and then click in the open reading frame.

“View sequence” gives you the nucleotide and amino acid sequence of the ORF.

“Blast” Blasts the ORF.

“Insert Gene” inserts the ORF. (see later slide for more on this)

# ORF Management in Manatee: gene adjustments

Refresh XML Search AsmbI Id: 7974 Database: gsp database: asmbI\_id: submit reset

| feat_name | end5    | end3    | role_id | ec_num | gene_sym | complete | com_name |
|-----------|---------|---------|---------|--------|----------|----------|----------|
|           | 2861863 | 2858294 |         |        |          |          |          |

Six Frame Options (on six frame clicks)  
View Sequence ☒ Blast ☐ Insert Gene ☐

Gene Options (on feat\_name clicks)  
View Sequence ☐ Annotate ORF ☒ Edit Start ☐ Blast ORF ☐ Merge Genes ☐ Delete Gene ☐

To make adjustments to existing genes in the database, click on the option you want to do, then click on the arrow for the gene of interest.

New pages will pop up with information specific for your request. (see later slides)

# ORF Management in Manatee: start edits

This page can be reached from the “Gene Options” menu on the Genome Viewer page or from the Gene Curation Page “Select Function pull-down by selecting “Edit Start Site

Purple text represents the gene of interest. Blue text represents other genes in the region. It is important not to introduce overlap with other genes when changing a start site. Often editing a start site will remove overlap between two genes. Occasionally an annotator may want to extend a start into an upstream gene and will find that the upstream gene in question is a small hypothetical with no homology to anything. In such a case the annotator should consider deleting the short hypothetical, since it becomes likely that it is not a real gene.

To edit a start, click on the start you want in the 6-frame representation. The new coordinate for the selected start site will appear in the “New End5” box. To save the change to the database, click “Submit”.

Region of 2856886 to 2858271

Green nucleotides are ribosome binding sites.

Purple bold amino acids are amino acids of your query gene.

Blue bold amino acids are amino acids of one of genes in the region (other than query gene).

Click on a new start to change the start site, then click submit to enter change to the database.

Feat\_name:

Name:

5' End:

3' End:

New End5:

| Start Edits |                                                               |            |
|-------------|---------------------------------------------------------------|------------|
| 2856725     | TTCCCGCTTCCAATCATGACGAACCTTGCAACTGCGACATTGAACACCCTTTTATTTTGT  | Nucleotide |
|             | F P L P I M T N L O L R H * T P F Y F C                       | Frame1     |
|             | S R F Q S * R T C N C D I E H P F I F V                       | Frame2     |
|             | P A S N H D E L A T A T L N T L L F L Y                       | Frame3     |
|             | K G S G I M V F K C S R C Q V G K * K Q                       | Frame4     |
|             | R G A E L * S S S A V A V N F V R K N K                       | Frame5     |
|             | E R K W D H R V O L O S M S C G K I K T                       | Frame6     |
| 2856785     | ATTTTACCTTGGCTAGGATAACCTCAGCCCTTAAACTGTCAACGCCAACCAAGTGATACAG | Nucleotide |
|             | I L P W L G * P Q P L N C Q R Q P V I O                       | Frame1     |
|             | F Y L G * D N L S P * T V N A N O * Y R                       | Frame2     |
|             | F T L A R I T S A L K L S T P T S D T G                       | Frame3     |
|             | I K G Q S P Y G * G K F Q * R W G T I C                       | Frame4     |
|             | Y K V K A L I V E A R L S D V G V L S V                       | Frame5     |
|             | N * R P * S L R L G * V T L A L W H Y L                       | Frame6     |
| 2856845     | GTTTACCACTGATTAATTTCAATCAACGCTGTGAGCTTTTATGCGCAATTTACTCGATT   | Nucleotide |
|             | V Y H * L I F N Q R C E L L C A I Y S I                       | Frame1     |
|             | F T T D * F S I N A V S F Y A O F T R F                       | Frame2     |
|             | L P L I N F Q S T L * A F M R N L L D F                       | Frame3     |
|             | T * W Q N I K L * R Q S S K H A I * E I                       | Frame4     |
|             | P K G S I L K * D V S H A K I R L K S S                       | Frame5     |
|             | N V V S * N E I L V A T L K * A C N V R N                     | Frame6     |
| 2856905     | TTGACTTTGATAGCGCCATATTTGGCACCTTATACCTCCATGACTCGTGCCTTCCTG     | Nucleotide |
|             | L T L I A P I F G T L I P P * L V H F L                       | Frame1     |
|             | * L * * R P Y L A P L Y L H D S C T S C                       | Frame2     |
|             | D F D S A H I W H P Y T S M T R A L P V                       | Frame3     |
|             | K V K I A G M N P V R I G G H S T C K R                       | Frame4     |
|             | K S K S L A W I Q C G * V E M V R A S G                       | Frame5     |
|             | Q S Q Y R G Y K A G K Y R W S E H V E Q                       | Frame6     |

# ORF Management in Manatee: other gene adjustments

## Insert

|                                            |              |
|--------------------------------------------|--------------|
| Insert gene for the following coordinates? |              |
| end5: 693924                               | end3: 693500 |
| <input type="button" value="YES"/>         |              |

In all cases you will be asked to confirm your request before it is carried out.

## Merge

**The gene shown will be preserved (for one coordinate and annotation info). The one you enter will be used to determine the new extended coordinates and then deleted. Be sure you want this entered gene deleted!**

|                                  |                                      |                                   |
|----------------------------------|--------------------------------------|-----------------------------------|
| Gene: ORF00755                   |                                      |                                   |
| Locus: TP0640                    | methyl-accepting chemotaxis protein  |                                   |
| gene_sym: mcp2-3                 | ec#:                                 | asmb1_id: 6333                    |
| end5: 699711                     | end3: 701552                         |                                   |
| Enter name of gene to merge with | <input type="text" value="gene id"/> | <input type="button" value="GO"/> |

## Delete

**Are You Sure You Want To Delete This Gene?**

|                  |                                     |                                    |
|------------------|-------------------------------------|------------------------------------|
| Gene: ORF00755   |                                     |                                    |
| Locus: TP0640    | methyl-accepting chemotaxis protein |                                    |
| gene_sym: mcp2-3 | ec#:                                | asmb1_id: 6333                     |
| end5: 699711     | end3: 701552                        | <input type="button" value="YES"/> |

# Annotation Checklist

- Look for HMM hits
  - evaluate what the HMMs are telling you - exact function? family membership? domain?
- Look at BER results
  - looking for proteins in the skim which are characterized (colored backgrounds)
  - many proteins are characterized but not marked so in our tables - may need to check proteins with white backgrounds to see if they are characterized
  - color coding does not indicate quality of match only that the match protein has been experimentally characterized
  - evaluate the alignment - what percent ID over what length? active sites? binding sites?
  - fill in characterized match accession number (by clicking on the accession in left column)
- Look at TMHMM, SignalP, Prosite, region, etc.
- Use multiple alignment (belvu link) and tree(tree icon link) as needed to differentiate function.
- Decide what you think the protein should be named
- Fill in appropriate fields for common name, gene symbol, EC#, comment.
- Decide what GO terms you need
  - find them on the page (HMMs, EC number, GO suggestions) or with the GO search tool
  - change/remove any IEA GO annotations
  - add GO evidence from HMMs, BER, Genome Properties, Prosite, etc.
- Review TIGR role and change as needed
- Check start site
  - look in BER and at the BER generated multiple alignment (belvu link)
  - adjust if necessary - using “edit start” function in pull down or in the Genome Viewer section
  - check start site box when finished curation
- Check “complete”, click “submit” and your done!

# External Manatee's limitations: things available in limited capacity

- Refresh Searches button - will not work, but you can submit sequences for re-searching, we hope to set up an automated pipeline for this, but the system is not in place yet. Currently there is an HMM search page on the CMR that can also be used.
- SignalP in the pull-down can work if you install it locally. Or you can go to the CBS site to run it on the fly:  
<http://www.cbs.dtu.dk/services>
- BER tree view - should work if you have Java on your machine, but may be tricky
- BER multiple alignment - you can view them with belvu if you are running on Linux (what we do), or you can try other multiple alignment tools, a possibility is:
  - <ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/>

# External Manatee's limitations: things not available

- consistency checks
- all frameshift scripts
- translation exceptions
- intergenic region analysis
- overlap analysis
- annotation status
- hypothetical protein list

# Acknowledgements

Heading up the effort:

Owen White

Jeremy Peterson

CMR:

Tanja Davidsen (Team leader)

Nikhat Zafar

Qi Yang

Prokaryotic Annotation:

Bill Nelson (Team leader)

Bob Dodson

Bob Deboy

Scott Durkin

Sean Daugherty

Ramana Madupu

Lauren Brinkac

Steven Sullivan

M.J. Rosovitz

Sagar Kothari

Susmita Shrivastava

HMMs:

Dan Haft

Jeremy Selengut

Bioinformatics Engineers:

Todd Creasy

Liwei Zhou

Sam Angiuoli

Charles Lu

Anup Mahurkar

And the many other TIGR employees present and past who have contributed to the development of these tools and to building the annotation protocols we use. Thanks also go to the funding agencies that support our work including NIH, NSF, and DOE.