Expert Assertions Through Community Annotation Jamborees

Lauren Brinkac
The J. Craig Venter Institute
What is Pathema?

- A NIAID Bioinformatics Resource Center designed to support bio-defense and infectious disease research.
- Pathema provides detailed curation and comparative analysis of six target pathogens:
  - **Category A priority pathogens:**
    - *Bacillus anthracis, Clostridium botulinum*
  - **Category B priority pathogens:**
    - *Burkholderia mallei, Burkholderia pseudomallei, Clostridium perfringens, Entamoeba histolytica*
Annotation. Currently we have "raw data" accumulating 100-fold faster than it is being annotated. GenBank and most other resources do not allow an "expert" on one gene to annotate that gene in the dozens of new complete genomes that come out each week, even if the expert was motivated to do so.

The static annotation in CMR is a real problem. It would be nice to have a constantly updated genome based on current research in the field. An example is what Fiona Brinkman has done with the Pseudomonas aeruginosa genome website. The B. mallei ATCC 23344 genome annotation has numerous errors that have been identified by researchers in the field and no one will update this information in the CMR or GenBank. If you want to do something that will actually push the field forward, you should work o

all display of gene annotations are easily link to pathway informations, pubmed publication literature report of each gene characterisation-validation, gene microarray result for transcription datas eg latest submission data by Dr Patrick Tan s data on gene expression in different growth period, proteomics-mass spec data for protein expression-immunoreactive proteins and localisation of protein if there is any reported immunohisto microscopy work
Annotation Protocols

Clostridium Protocols

Submit Your Own Protocol

Standard Operational Procedures
Biosafety Protocols
Detection Protocols
Epitope Protocols
Reagent Links
Safety Requirements

Standard Operational Procedures

This following is completed documentation of many of JCVI's standard operating procedures (SOPs) for gene model and functional curation. Each SOP outlines the purpose and scope of each process or procedure, all requirements necessary to carry out each procedure, a detailed description of the process or analysis, and all measures utilized to ensure that data and data types generated are consistent, current, and maintain specific quality requirements.

- **Gene Model Curation**
  - Gene Prediction: Outlines the process for locating and predicting coding and non-coding genome features.
  - Analysis and Curation of Short Gene Models: Describes the process used by curators for evaluating potential false positives generated by Glimmer.
  - Creation of a Pseudomolecule: Describes the process of creating a "pseudomolecule" for an unfinished genomic sequence.
  - Annotation mapping: Describes the procedure of mapping gene models and annotation from a reference genome to a related molecule.
  - Homology Searches: Outlines all of our pre-computed homology searches run to generate evidence for functional annotation.

- **Functional Curation**
  - Functional Automated Annotation: Defines our automated annotation program that generates putative functional annotation to each gene model.
  - Functional Manual Curation: Documents the layered evidences based approach used by curators for assigning descriptive functional annotation to each gene model.
  - Start Site Curation: Describes the criteria used by curators when evaluating and editing the initiation codon for each gene model.
  - Frameshift Edit and Analysis: Describes the procedure used by curators for evaluating and editing potential frame shifts, point mutations, and sequence ambiguities.
  - Overlap Analysis and Curation: Defines the criteria used by curators when reviewing overlapping regions of gene models.
  - Supporting Documentation
    - Naming Convention Guidelines

Gene Model Curation

Functional Curation

Homology Searches
Naming Convention Guidelines

JCVI Gene Naming and Annotation Conventions

All genes sequenced at TIGR are initially assigned annotation through an automated process. Names and functional annotation are then manually curated. As direct experimental evidence rarely exists for each gene in a sequenced organism, name assignments are usually based on sequence similarity. Therefore, all TIGR name assignments should be regarded as provisional. We strive to annotate each gene with as much information as we can confidently impart, but are also wary of assigning too much from sequence similarity. We prefer to err on the side of caution and have devised a nomenclature scheme that reflects our degree of confidence in a particular assignment.

We encourage feedback from the community to help identify errors or to provide suggestions to improve the annotation of our genes.

Evidence Types

Levels of Database Match

Levels of Database Match

Decisions for annotation include a minimum common name, role category, and Gene Ontology (GO) "function" and "process" terms for each gene, and may also include a gene symbol, an Enzyme Commission number, and public comments. Each gene is assigned as many descriptors as are relevant. In the course of reviewing data we have developed the following criteria regarding assignments.

Specific functions - indicated by a specific name and gene symbol

- The protein translation has a good database match to a protein that its function and process have been experimentally characterized. Both name and multiple protein sequence alignments reveal a high degree of identity/similarity (typically >80% identity) along the entire length of the protein. There may be one or more full-length matches to a high specificity (i.e., "equivalent ontology" type) HMM. Active sites, substrate or catalytic binding sites, or motifs that are characteristic of the protein should be conserved. Strong conservation of gene content (e.g. operon structure) is also taken as evidence for certain function. For genes with a certain function we use the most widely recognized name and gene symbol. Highly specific GO function and process terms are used if available. Enzyme Commission numbers are annotated with their EC numbers. The EC number of the enzyme domain may also be used for clarity. It may seem more informative than its common name.

- When single or more lines of evidence are weak, but most of the data agree, we conclude the gene is likely performing the function the name implies, and the name is therefore a "putative" or "putative" in the name.

- If one or more lines of evidence are weak, but most of the data agree, we conclude the gene is likely performing the function the name implies, and the name is preceded with "putative". In such cases, the percent identity (e.g. 30-35% identity) of the HMM score (e.g.between the predicted and query database) is not quite high enough to impart certainty. GO terms are more general than for specific gene functions of certain function, and with few exceptions, gene symbols are not used. For an enzyme with a specific function, specific protein family and/or domain names may be used.

- When there is strong evidence for conflicting evidence, we consider the evidence indicated by the common name to be unlikely. We add "homolog" to the common name. Such assignments can arise from two situations: Two situations. First, the gene homolog sequence is very strong, but unlike the query protein, we do not believe the query protein has the same function as the database match. In this case, because some critical elements of protein are absent (e.g., non-essential catalytic residues in an enzyme), or because the function is predicted to assist in this particular organism (e.g., photosynthetic context, CNS), the second situation, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessory domain, or the accessi
Pathema Gene Page

Current Annotation

Curation Status

Evidence

Annotation

JCVI Annotation Display: EHL_142000

Locus Name: EHL_142000
Putative identification: histone acetyltransferase, putative
Coordinates: 79753-78509
DNA Molecule Name: 110 1688137485 135674 bp Entamoeba histolytica HM1-IMSS
Gene length: 1244 nt
Protein length: 415 aa
Molecular Weight: 48490.29
pl: 7.2506
Gene Ontology (GO) Role Category: GO:0003723: molecular_function, RNA binding
Gene Ontology (GO) Role Category: GO:0004402: molecular_function, histone acetyltransferase activity
Curation Status: JCVI manual curation complete
Community Annotation: Click here to submit new annotation for this gene
Web Community Submission

Registration

Update Annotation

Provide Inference and Evidence

Curator
Direct Assay
Expression Pattern
Genetic Interaction
Mutant Phenotype
Physical Interaction
Community Annotation

JCVI Annotation Display: EHL_142000

- Locus Name: EHL_142000
- Putative identification: histone acetyltransferase, putative
- Coordinates: 79753-78500
- DNA Molecule Name: 11016281837485 135674 bp 6 Entamoeba histolytica HM_1.EMSS
- Gene length: 1244 nt
- Protein length: 415 aa
- Molecular weight: 48490.29
- pl: 7.2305

- Gene Ontology (GO) Role Category:
  - GO:0003723: molecular_function, RNA binding
  - GO:0004402: molecular_function, histone acetyltransferase activity

- Curation Status: JCVI manual curation complete

External Curator Information:
- Name: Girja Ramakrishnan
- Institute: University of Virginia
- Proposed Gene Information:
  - 5 Prime End: 79753
  - 3 Prime End: 78500
- Protein Name: histone acetyltransferase, putative
- Gene Symbol: EHYST
- EC Number: 2.3.1.48

Evidence Information:
- Inferred From: Inferred from Direct Assay (IDA)
- PubMed Reference: 15565732
- Other Evidence: Transcription analysis by RT-PCR
- Commons: Enzymatic activity was determined of recombinant protein expressed in E. coli.
Workshops and Training

- **On JCVI Campus**
  - 2 Days Hands On
  - Live/Archived Webinar Broadcast
  - 4 Workshops Conducted, one per clade

- **Organism Specific Meetings**
  - 4 Hours Hands On
  - 5 Workshops Conducted

- 148 Researchers Attended
- 123 Community Registration
Web Community Submissions

Totals To Date:
4 Annotation Updates
  - 1 gene structure
  - 1 protein name
  - 3 gene symbols
  - 1 enzyme number
  - 4 literature references
Advantages and Disadvantages

**Advantages**
- Value added to annotation
- Convenient & accessible for the researcher
- Minimal effort needed
- Low cost

**Disadvantages**
- Training not required
  - Lack of adherence to standards
- Incentive and intimidation
Annotation Jamboree Goals

• **Incorporate** expert annotation data into existing Pathema genome submissions.
• **Update** existing annotation based on expert assertions.
• **Tag** annotation with updated experimental references.
• **Provide** the community with the opportunity to become familiar with JCVI annotation procedures.
Annotation Jamboree Details

- 3 day Jamborees held at JCVI
- First day teaches our annotation methodologies and Manatee
- Follow up days allow the community to annotate with aid from our expert curators and present latest research on gene sets
- Data dissemination and participant recognition within Pathema and GenBank

Excellent participation and feedback from our participants
Evidence Standards

- **phosphomethylpyrimidine kinase**
  - Function was **inferred from direct assay** using the accession SP|P76422 with reference PMID: 10075431
  - Process was **inferred from direct assay** using the accession SP|P76422 with reference PMID: 10075431

- **putative hypoxanthine phosphoribosyltransferase**
  - Function was **inferred from sequence similarity** using the accession TIGR01203
  - Process was **inferred from mutant phenotype** using the accession SP|Q02522 with reference PMID: 1465108
**MGAT: Multi-Gene Annotation Tool**

<table>
<thead>
<tr>
<th>Ref</th>
<th>C</th>
<th>Start</th>
<th>Edit</th>
<th>Gene</th>
<th>GO Information</th>
<th>Gene Name</th>
<th>Symbol</th>
<th>EC#</th>
<th>Length</th>
<th>Role</th>
<th>BER match</th>
<th>GenProp</th>
<th>HMM</th>
<th>HMM Prop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gene</td>
<td>Name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Multiple Gene Annotation Tool**

Log into [php666] as [jrvona]

[Home] [View Tree Aligment] [Non-Redundant Blast] [Refresh Searches]

---

**J. Craig Venter INSTITUTE**
### Paralogous Family Annotation

<table>
<thead>
<tr>
<th>Genes/Proteins</th>
<th>GO ID</th>
<th>Select Action</th>
<th>Show Option</th>
<th>Introns Options</th>
<th>Full Length</th>
<th>Show No HMMs</th>
<th>HMM</th>
<th>Para Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetyltransferase, GNAT family</td>
<td>GO:0008152</td>
<td>[Select Action]</td>
<td>Show all</td>
<td>Show no HMMs</td>
<td></td>
<td></td>
<td>PP00583</td>
<td>Para domain</td>
</tr>
<tr>
<td>Acetyltransferase, GNAT family</td>
<td>GO:0008152</td>
<td></td>
<td>Show all</td>
<td>Show no HMMs</td>
<td></td>
<td></td>
<td>PP00583</td>
<td>Para domain</td>
</tr>
<tr>
<td>Acetyltransferase, GNAT family</td>
<td>GO:0008152</td>
<td></td>
<td>Show all</td>
<td>Show no HMMs</td>
<td></td>
<td></td>
<td>PP00583</td>
<td>Para domain</td>
</tr>
<tr>
<td>Acetyltransferase, GNAT family</td>
<td>GO:0008152</td>
<td></td>
<td>Show all</td>
<td>Show no HMMs</td>
<td></td>
<td></td>
<td>PP00583</td>
<td>Para domain</td>
</tr>
<tr>
<td>Acetyltransferase, GNAT family</td>
<td>GO:0008152</td>
<td></td>
<td>Show all</td>
<td>Show no HMMs</td>
<td></td>
<td></td>
<td>PP00583</td>
<td>Para domain</td>
</tr>
<tr>
<td>Acetyltransferase, GNAT family</td>
<td>GO:0008152</td>
<td></td>
<td>Show all</td>
<td>Show no HMMs</td>
<td></td>
<td></td>
<td>PP00583</td>
<td>Para domain</td>
</tr>
<tr>
<td>Acetyltransferase, GNAT family</td>
<td>GO:0008152</td>
<td></td>
<td>Show all</td>
<td>Show no HMMs</td>
<td></td>
<td></td>
<td>PP00583</td>
<td>Para domain</td>
</tr>
<tr>
<td>Acetyltransferase, GNAT family</td>
<td>GO:0008152</td>
<td></td>
<td>Show all</td>
<td>Show no HMMs</td>
<td></td>
<td></td>
<td>PP00583</td>
<td>Para domain</td>
</tr>
</tbody>
</table>
# Annotation Jamboree Results

<table>
<thead>
<tr>
<th>Jamboree Statistics</th>
<th>Entamoeba</th>
<th>Burkholderia</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Participants</strong></td>
<td>14</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>Community Researchers</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>JCVI Analysts</td>
<td>8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td><strong>Curated Functional Assignments</strong></td>
<td>1022</td>
<td>543</td>
<td>1565</td>
</tr>
<tr>
<td>Assignments made by community researchers</td>
<td>595</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Assignments made by JCVI based on community contributions</td>
<td>n/a</td>
<td>404</td>
<td></td>
</tr>
<tr>
<td>Assignments made by JCVI</td>
<td>427</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td><strong>GO Terms Assigned</strong></td>
<td>2019</td>
<td>1480</td>
<td>3499</td>
</tr>
<tr>
<td>Assignments made by community researchers</td>
<td>123</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Assignments made by JCVI based on community contributions</td>
<td>500</td>
<td>1198</td>
<td></td>
</tr>
<tr>
<td>Assignments made by JCVI</td>
<td>1396</td>
<td>282</td>
<td></td>
</tr>
<tr>
<td><strong>Paralogous Families Curated</strong></td>
<td>120</td>
<td>n/a</td>
<td>120</td>
</tr>
<tr>
<td><strong>Gene Structures Edited</strong></td>
<td>43</td>
<td>86</td>
<td>129</td>
</tr>
<tr>
<td>Assignments made by JCVI based on community contributions</td>
<td>43</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>Assignments made by JCVI</td>
<td>0</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td><strong>Experimental References Added</strong></td>
<td>285</td>
<td>11</td>
<td>296</td>
</tr>
<tr>
<td><strong>Number of Genome Projects Updated</strong></td>
<td>1</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td><strong>Number of Post Collaborations Established</strong></td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
## Burkholderia Jamboree

A three-day, hands-on Burkholderia Annotation Jamboree was conducted in an effort to train community experts in the Burkholderia field on JCVI’s internal tools for structural and functional annotation. The jamboree was held on the JCVI campus, Sept 24, 2008 - Sept 26, 2008. Training consisted of a series of lectures describing JCVI’s probiotic strain methodologies and functional annotation tools. Additionally, community researchers participated in a series of interactive hands-on annotation training exercises. After completing the training portion of the jamboree, community researchers subsequently contributed to the manual curation of *Burkholderia* genome projects.

### Jamboree Statistics

<table>
<thead>
<tr>
<th>Total Jamboree Contributions</th>
<th>Participants</th>
<th>Functional Names Assigned</th>
<th>GO Terms Assigned</th>
<th>Gene Structural Curation</th>
<th>Literature References</th>
<th>Burkholderia genomes updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>9</td>
<td>502</td>
<td>1460</td>
<td>85</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

#### Participants

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Burkholderia Researchers</td>
<td>6</td>
</tr>
<tr>
<td>JCVI Analysts</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Functional Names

| Assignments made by Burkholderia Researchers | 41           |
| Assignments made by JCVI based on community contributions | 404          |
| Assignments made by JCVI | 98           |

#### GO Terms

| Assignments made by JCVI based on community contributions | 1190         |
| Assignments made by JCVI | 282          |

#### Gene Structural Curation

| Edited gene structures by JCVI based on community contributions | 69           |
| Edits made by JCVI | 17            |

#### Literature References

| Functional Assignments with associated experimental references | 11           |
JCVI Annotation Display: BURPS1106B_A1276

Comment: This functional annotation is a direct result of community participation in a Pathema-Burkholders annotation jamboree.

Role Categories:
- Cellular Role Category: Central intermediary metabolism: Other
- Gene Ontology (GO) Role Category: GO:0008152: biological_process, metabolic process
- Gene Ontology (GO) Role Category: GO:0050538: molecular_function, N-carbamoyl-L-amino-acid hydrolase activity

Attributes:
- Coordinates: 1201014-1289734
- Gene length: 1281 nt
- Protein length: 426 aa
- Molecular Weight: 44990.38 kDa
- pI: 5.844

<table>
<thead>
<tr>
<th>GO Term</th>
<th>Type</th>
<th>GO Term Name</th>
<th>GO Term Definition</th>
<th>Evidence Code</th>
<th>With</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0008152</td>
<td>biological_process</td>
<td>metabolic process</td>
<td>The chemical reactions and pathways, including...</td>
<td>inferred from sequence similarity</td>
<td>GENBANK:CA69909.1</td>
<td>GO_REF:0006012</td>
</tr>
<tr>
<td>GO:0056538</td>
<td>molecular_function</td>
<td>N-carbamoyl-L-amino-acid hydrolase activity</td>
<td>Catalysis of the reaction...</td>
<td>inferred from sequence similarity</td>
<td>GENBANK:CA69909.1</td>
<td>GO_REF:0006012</td>
</tr>
</tbody>
</table>
- Contributors become authors on submission updates
- Annotation jamboree & participating institutions acknowledged in comments
- db_xrefs to gene page established

LOCUS  ABM89094  176 aa  linear  DCT 03-APR-2009
DEFINITION  shock protein HaLV, ATP-dependent protosase subunit HaLV
[Burkholderia pseudomallei 1106a].
ACCESSION  ABM89094
VERSION  ABM89094.2  GI:210148308
DBSOURCE  accession CP000712.1
KEYWORDS  
SOURCE  Burkholderia pseudomallei 1106a
ORGANISM  Burkholderia pseudomallei 1106a
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Burkholderia; pseudomallei group.
REFERENCE  1  (residues 1 to 178)
AUTHORS  Harkin,B.M., Brinkac,L.K., Brown,K.A., Hung,G.C., Tuanyok,A.,
Zhang,B. and Nierman,W.C.
TITLE  Revision of Burkholderia pseudomallei 1106a
JOURNAL  Unpublished
REFERENCE  2  (residues 1 to 178)
AUTHORS  DeShazer,D., Woods,D.E. and Nierman,W.C.
TITLE  Direct Submission
JOURNAL  Submitted (15-FEB-2007) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REFERENCE  3  (residues 1 to 178)
AUTHORS  Harkin,B.M.
TITLE  Direct Submission
JOURNAL  Submitted (30-OCT-2008) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REMARK  Protein update by submittter
REFERENCE  4  (residues 1 to 178)
AUTHORS  Harkin,B.M.
TITLE  Direct Submission
JOURNAL  Submitted (19-NOV-2008) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REMARK  Protein update by submittter
COMMENT  On Oct 31, 2008 this sequence version replaced gi:126225554.
Source DNA is available from BRCresources
(http://www.brcresources.org/).

This functional annotation is the result of community participation in the Pathema-Burkholderia annotation jamboree held at the J. Craig Venter Institute, September 24-26, 2008
(http://pathema.jcvi.org/). Participating institutions for the genome revision: The J. Craig Venter Institute, University of Texas at Austin, Imperial College London, Armed Forces Institute of Pathology, and Northern Arizona University.
Long Term Benefits

• **Continued collaborations**
  o 4 Community Researchers

• **Expert assertions**
  o 20 additional assertions / 2 months
  o Curation of genomic islands and SSRs

• **Dissemination of annotation standards**

• **Custom analysis and data generation**
Advantages and Disadvantages

• **Advantages**
  - Value added to annotation
  - In depth hands on training
    - Adherence to standards
  - High incentive and lower intimidation
  - Long term benefits

• **Disadvantages**
  - Large cost
  - Time commitment
Next Annotation Jamboree

• Pathema - *Bacillus/Clostridium* Annotation Jamboree
  - **When:** May 13-15, 2009
  - **Cost:** Free
  - **Update annotation with expert assertions**
    - 19 *Bacillus anthracis*
    - 11 *Bacillus cereus*
    - 12 *Clostridium botulinum*
    - 9 *Clostridium perfringens*

• **More information:** [http://pathema.jcvi.org](http://pathema.jcvi.org)
Acknowledgments

PI: Granger Sutton

Informatics Engineers
Tanja Davidsen (manager)
Erin Beck
Kevin Galinsky
Jason Inman
Alex Richter
Bob Montgomery

Informatics Analysts
Scott Durkin (manager)
Ramana Madupu
Bob Dodson
Derek Harkins
Susmita Shrivastava
Lis Caler

All others past & present who contributed to the development and data analyses of this resource.

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN266200400038C