Taxonomy Quick Guide
Objective

• Help ease the process of providing appropriate names to organisms to be sequenced by the JGI and identify potential errors in naming early.

• This guide provides an overview of organism taxonomy for isolates and single cells

• We will describe the process for identification of inconsistencies between sequencing project names and organism names and how to address these issues.
Taxonomy is the science of defining groups of biological organisms on the basis of shared characteristics.

**Yeast**
- Eukarya
- Fungi
- Ascomycota
- Saccharomycetes
- Saccharomycetales
- Saccharomyces
- S. cerevisiae

**E. coli**
- Bacteria
- Prokaryotes
- Proteobacteria
- Gammaproteobacteria
- Enterobacteriales
- Enterobacteriaceae
- Escherichia
- E. coli
Verifying Isolate Projects

• Another example:
  ➢ The project name field should be the same as the organism name

ITS Project ID: 1104880:
Project Name: Actinobacterium Strain MWH-Ta3
Organism name: *Erythroluna alpina* MWH-Ta3

<table>
<thead>
<tr>
<th>Life</th>
<th>Project Name</th>
<th>Organism Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
<td>Actinobacterium</td>
<td></td>
</tr>
<tr>
<td>Kingdom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phylum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genus</td>
<td></td>
<td><em>Erythroluna taiensis</em></td>
</tr>
<tr>
<td>Species</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain</td>
<td>Strain MWH-Ta3</td>
<td>MWH-Ta3</td>
</tr>
</tbody>
</table>
More things to look out for....

• **Genus/species field is left blank**
  – Possible cause: Investigator is sequencing a new organism and is unsure of the identity
  – Resolution:
    • Investigator must provide some information for higher taxonomic levels (e.g. Gammaproteobacteria sp. 123)
    
    - In this case, the organism and project name will be “Gammaproteobacteria sp. 123”. The GOLD convention is to leave the “genus” field blank since this is not well defined, populate the entire name in the “species” field and populate the “strain” field
    • If the investigator is proposing a new genus, we accept it without a publication with the understanding that if/when the project is registered with NCBI, they may not accept the new genus without an associated publication and suggest alternative names until a publication is available describing the genus

• **Strain field is left blank (this is a RED FLAG)**
  – Possible cause: Investigator is sequencing a new organism and is unsure of the identity
  – Resolution:
    • For bacterial and fungal projects, the strain is used to distinguish between two organisms within the same genus/species and cannot be left blank. The investigator must specify a strain ID.
    • If the organism has a repository ID (culture collection ID), that can be used as the strain ID
Repository ID (or Culture Collection ID): Culture collection centers are repositories for organisms & reference materials. Examples include:

- **DSMZ**
- **ATCC**
- **NCTC** (National Collection of Type Cultures) operated by Public Health England
- **NRRL**
Verifying Isolate Names

- If you have doubts about a name, you can consult StrainInfo

http://www.straininfo.net/

Taxon Passport
*Escherichia coli*

<table>
<thead>
<tr>
<th>overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>species</td>
</tr>
<tr>
<td>parent taxon</td>
</tr>
<tr>
<td>16S rRNA gene</td>
</tr>
<tr>
<td>external links</td>
</tr>
</tbody>
</table>
Verifying Isolate Names

• You can also consult the NCBI Taxonomy Browser


Escherichia coli DSM 30083 = JCM 1649 = ATCC 11775

Taxonomy ID: 866789
Inherited blast name: enterobacteria
Rank: no rank
Genetic code: Translation table 11 (Bacterial, Archaeal and Plant Plastid)
Other names:
synonym: Escherichia coli JCM 1649
synonym: Escherichia coli DSM 30083
synonym: Escherichia coli CIP 54.8
synonym: Escherichia coli ATCC 11775

Lineage (full)
  cellular organisms; Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia; Escherichia coli
Verifying Isolate Names

- Another resource is DSMZ

https://www.dsmz.de/bacterial-diversity/helpful-links-regarding-taxonomy.html
Key Points

- The organism name should be a part of the project name (at least the stem)
- Both genus and species fields cannot be empty. The investigator must provide some taxonomic identifier for the sample to be sequenced.
- Strain fields for isolates cannot be empty
- If you have questions about naming, consult your JGI project manager for assistance prior to submission of project metadata
Sample and metadata submission information can be found at:

http://jgi.doe.gov/user-program-info/pmo-overview/project-materials-submission-overview/