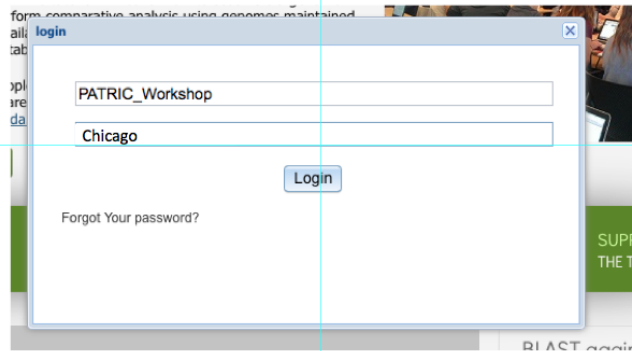


# Annotating a Genome in PATRIC

The following step-by-step workflow is intended to help you learn how to navigate the new PATRIC workspace environment in order to annotate and browse your genome on the PATRIC website.

## Step 1. Go to the PATRIC website and log on

- [www.patricbrc.org](http://www.patricbrc.org)
- login: workshopdemo
- Password: patric123

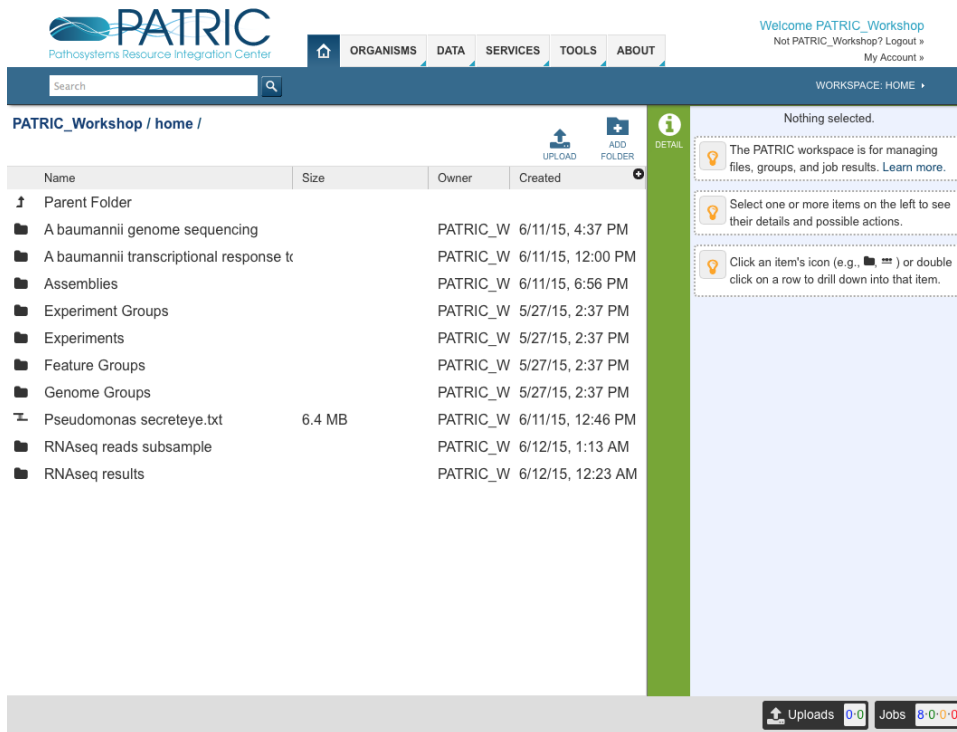


## Step 2. Let's familiarize ourselves with the PATRIC workspace

- Click on Workspace: Home in the top right corner to go to the PATRIC workspace



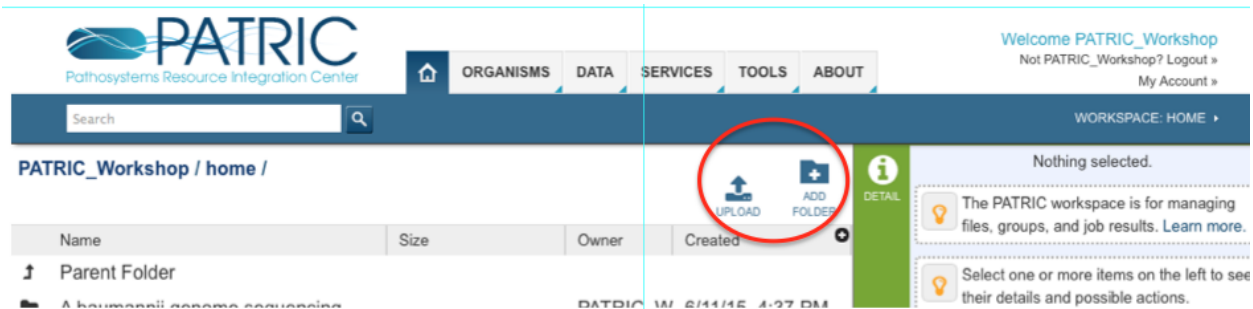
- The workspace is a series of folders that you can make for your projects



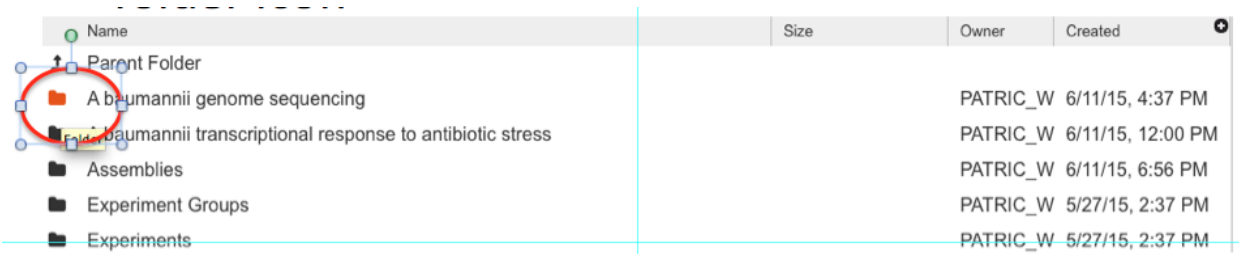
- The right hand panel shows links for information on how to use the workspace
- This can be displayed or hidden by clicking on the "DETAIL" button



- To create a folder or upload to a folder click on the buttons in the top right corner of the main panel



- You can navigate into a folder by clicking on the folder icon

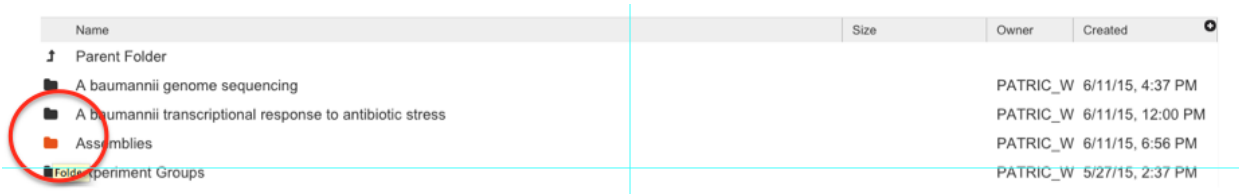


- The current working directory will be shown at the top
- The "Parent Folder" Icon is your "Back" button



### Step 3. Download two *Acinetobacter baumannii* genomes to your computer

- Navigate to the "Assemblies" folder
- Click on the first genome, A. baumannii 1000160
- Click on the SRR1033639.auto assembly
- Download it to your computer
- Note that you may need to turn off the pop-up blocker on your browser



Name	Size	Owner	Created
Parent Folder			
A baumannii 1000160		PATRIC_W	6/11/15, 7:04 PM
Folder baumannii 1592897		PATRIC_W	6/11/15, 6:58 PM


Name	Size	Owner	Created
Parent Folder			
SRR1033693.auto	4.4 kB	PATRIC_W	6/11/15, 8:31 PM

**PATRIC\_Workshop / home / Assemblies / A baumannii 1000160 / SRR1033693.auto**

Genome Assembly Job Result

<b>Start time</b>	6/11/15, 7:05 PM
<b>End time</b>	6/11/15, 8:31 PM
<b>Run time</b>	1h26m3607s
<b>Parameters</b>	<pre>{   "output_file": "SRR1033693.auto",   "output_path": "/PATRIC_Workshop@patricbrc.org/home/Assemblies/A baumannii 1000160",   "paired_end_libs": [     {       "read2": "/PATRIC_Workshop@patricbrc.org/home/A baumannii genome sequencing/Acinetobacter baumannii 1000160/SRR1033693_2.fastq.gz",       "read1": "/PATRIC_Workshop@patricbrc.org/home/A baumannii genome sequencing/Acinetobacter baumannii 1000160/SRR1033693_1.fastq.gz",       "interleaved": "false"     }   ],   "recipe": "auto",   "reference_assembly": "" }</pre>

Result Files

Filename	Type	File size
 SRR1033693.auto.contigs	contigs	4.0 MB

*Make sure that the pop-up blocker is turned off on your browser*

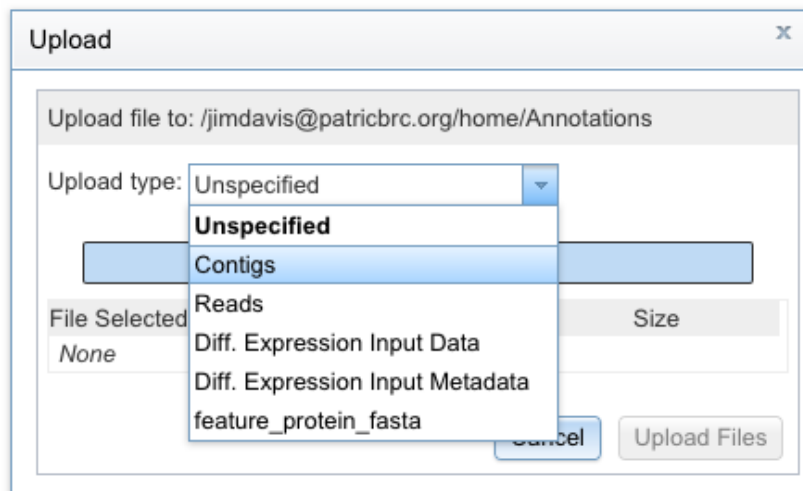
- Repeat this for *A. baumannii* 1592797.
- Select the SRR1019284.auto assembly

**Step 4. Let's annotate our *A. baumannii* genomes in our own workspace**

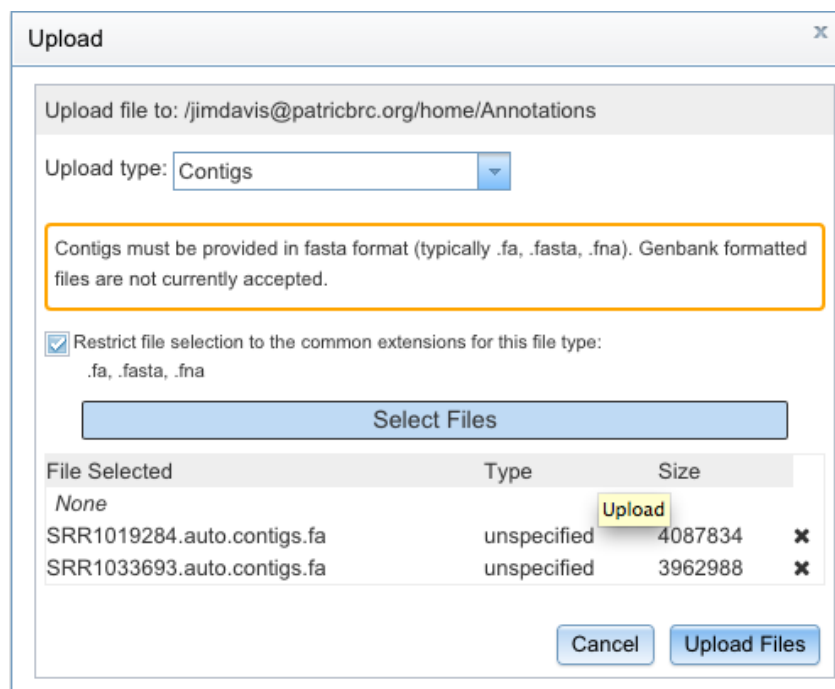
- Locate both files on your computer
- Add the file extension .fa to both files
- Log out of the "Patric\_Workshop" workspace
- Log into your personal workspace
  - If you do not have an account, please create one
- In your workspace create a new folder called "Annotations"
- Navigate to the Annotations folder
- Click the upload icon

Name	Size	Owner	Created
↑ Parent Folder			

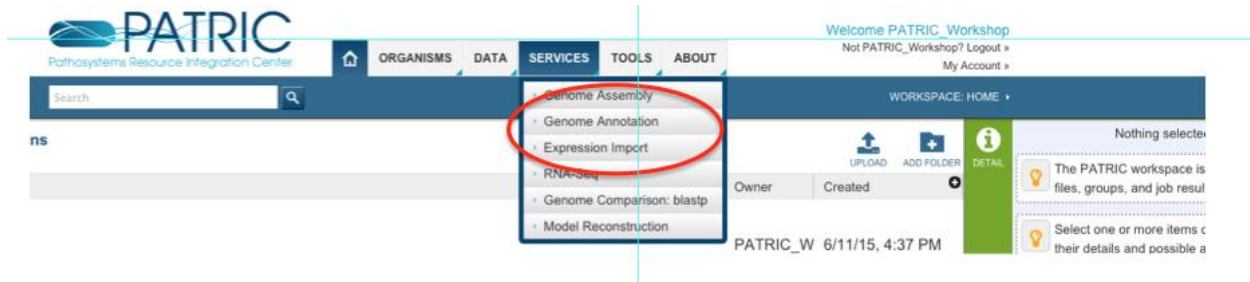
- Select "Contigs" as the file type



- Click "Select Files", find the two contig files and upload them



- Under the "Services" tab at the top of the page, select "Genome Annotation"



- First select the Annotations folder
- Then select one of the contig files

Genome Annotation  
Annotates genomes using RASTik.

**Parameters**

CONTIGS

DOMAIN  
 Bacteria

ORGANISM NAME  
 e.g. Bacillus Cereus

TAXONOMY ID

GENETIC CODE  
 11 (Archaea & most Bacteria)

OUTPUT FOLDER

OUTPUT NAME  
 Output Name

Choose or Upload a Workspace Object

Folder: fjimdavis@patricbrc.org/home/Annotations  
 Selection: SRR1019284.auto.contigs.fa

Name	Created
↑ Parent Folder	
☒ SRR1019284.auto.contigs.fa	6/12/15, 11:30 AM
☒ SRR1033693.auto.contigs.fa	6/12/15, 11:30 AM

Show files with an unspecified type

- Select SRR1033693.auto
- Fill out the form as shown below

## Genome Annotation

Annotates genomes using RASTtk.

**Parameters** ⓘ

CONTIGS  
SRR1033693.auto.contigs

DOMAIN  
Bacteria

ORGANISM NAME  
Acinetobacter baumannii 1000160

TAXONOMY ID  
1310800

GENETIC CODE  
11 (Archaea & most Bacteria)

OUTPUT FOLDER  
Annotations

OUTPUT NAME  
Acinetobacter baumannii 1000160

Reset

Annotate

- Click "Annotate"
- Do the same for the other genome

**Parameters** ⓘ

CONTIGS  
SRR1019284.auto.contigs

DOMAIN  
Bacteria

ORGANISM NAME  
Acinetobacter baumannii 1592897

TAXONOMY ID  
1310696

GENETIC CODE  
11 (Archaea & most Bacteria)

OUTPUT FOLDER  
Annotations

OUTPUT NAME  
Acinetobacter baumannii 1592897

Reset

Annotate

- You can see the job status in the bottom right
- Clicking here will show the completed and queued jobs

Status	Submit	App	Output Name	Start	Completed
queued	6/12/15, 11:36 AM	GenomeAnnotation	Acinetobacter baumannii 1592897		
queued	6/12/15, 11:33 AM	GenomeAnnotation	Acinetobacter baumannii 1000160		
● failed	6/10/15, 3:41 PM	GenomeAnnotation	83333.1.anno	6/10/15, 3:41 PM	
● completed	6/3/15, 4:37 PM	GenomeAnnotation	E_Coli_TEST	6/3/15, 4:37 PM	6/3/15, 4:42 PM

## What is happening during our annotation job?

We are using a version of RAST that has been customized for PATRIC

- Calling rRNAs (16S, 23S, 5s)
- Calling tRNAs with tRNAscanSE
- Searching for repeat regions
- Finding special proteins
  - Selenoproteins
  - Pyrrolysylproteins
- If it is a Streptococcus genome, it will find special repeats
- Calling CRISPRs
  - clustered regularly interspaced short palindromic repeats
- Calling protein-encoding genes
  - Prodigal
  - Glimmer
- Assigning function
  - First attempt: annotates against CoreSEED
  - Second attempt: annotates against FIGFams
  - Third attempt: BLAST against close relatives
- Overlapping genes are resolved
- Annotates matches to ARDB, CARD, virulence factors, human homologs
- Assigns proteins to FigFams
- Finds closest neighbors

## Step 5. Let's look at our annotation output

- If you click on the annotation job you will see job details:





















## Genome Annotation Job Result

<b>Genome</b>	Organism (Acinetobacter baumannii 1000160), Domain (Bacteria), Annotation ID (1310800.5), Feature count (3864)
<b>Start time</b>	6/12/15, 1:43 PM
<b>End time</b>	6/12/15, 2:31 PM
<b>Run time</b>	47m32s
<b>Parameters</b>	{ "scientific_name": "Acinetobacter baumannii 1000160", "domain": "Bacteria", "contigs": "/PATRIC_Workshop@patricbrc.org/home/Assemblies/A baumannii 1000160/SRR1033693.auto/SRR1033693.auto.contigs", "output_file": "Acinetobacter baumannii 1000160", "output_path": "/PATRIC_Workshop@patricbrc.org/home/Annotations", "taxonomy_id": "1310800", "code": "11" }

- And a set of files that you can download:

### Result Files

Filename	Type	File size
 Acinetobacter baumannii 1000160.gff	gff	593.3 kB
 Acinetobacter baumannii 1000160.txt	string	5.3 MB
 Acinetobacter baumannii 1000160.embl	embl	8.4 MB
 Acinetobacter baumannii 1000160.tar.gz	string	2.2 MB
 Acinetobacter baumannii 1000160.feature_protein.fasta	feature_protein_fasta	1.6 MB
 Acinetobacter baumannii 1000160.merged.gb	genbank_file	8.1 MB
 Acinetobacter baumannii 1000160.gb	genbank_file	8.0 MB
 Acinetobacter baumannii 1000160.xls	string	5.6 MB
 Acinetobacter baumannii 1000160.features.txt	feature_table	573.1 kB
 Acinetobacter baumannii 1000160.genome	genome	10.6 MB
 Acinetobacter baumannii 1000160.contigs.fasta	contigs	4.0 MB
 Acinetobacter baumannii 1000160.feature_dna.fasta	feature_dna_fasta	3.9 MB
 sp_gene.json	json	104.1 kB
 load_files	folder	0 B
 genome.json	json	1.1 kB
 genome_feature.json	json	8.7 MB
 genome_sequence.json	json	3.9 MB
 pathway.json	json	2.5 MB

- If you click on the icon button, you will download the files to your folders.
- Typically, you will want the genbank or fasta formatted file.
- The .json and .genome files are java script files that are used by our system

 Acinetobacter baumannii 1000160.xls	string	5.6 MB
 Acinetobacter baumannii 1000160.feature_dna.fasta	feature_dna_fasta	3.9 MB
 Acinetobacter baumannii 1000160.feature_protein.fasta	feature_protein_fasta	1.6 MB
 Acinetobacter baumannii 1000160.gff	gff	593.3 kB
 load_files	folder	0 B
 genome.json	json	1.1 kB

- The download box may look different if you are using a PC



- Using your favorite text editor you can see the proteins in fasta format (in this case).

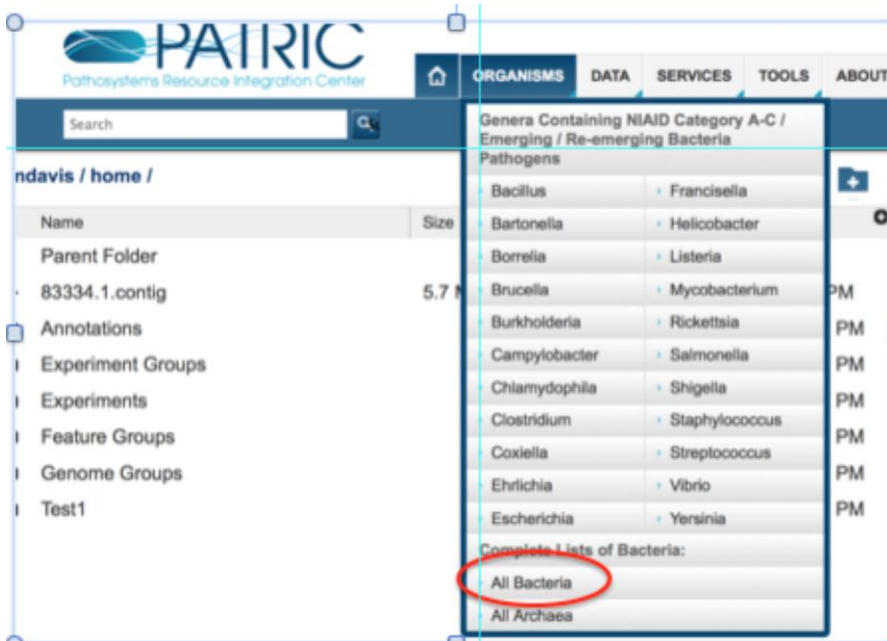
```

Acinetobacter-1
>fig|1310800.6.peg.1 High-affinity choline uptake protein BetT [Acinetobacter 1310800.6]
MPSKPS5QFANVNPVFLSTVIIAIFLAIVIVAPSSFELLTQQLKQWITDSFSWFYVLS
VAVFLILLIYIACSASGKIKLGPDHSQPEYSNGSWFAMLF TAGMGIGLMFFGVAEPMHY
VSPSPGDPEVESQAQALRITFFHWGLHAWAIYAVVGLALAYFAYRHNLPKTRSAIYPI
IGKKIYGPIDGIDTFATIGTVFVGVATSLGFGVTQINSGLHYLFGVEQSSYIQVGLIIFV
SILASMSVFLGLDKGVKRLSELNLILALVLLAFVFIAGPSIYLLQTTIQNAGSYFSNLF
MTFNLYAYQPNGWIGGWIMYAWWISWSPFVGMFIARVSRGRTIREFIVGVLLIPTGFT
IWMGFLGNAALFSIMQEHNVTLQAVQNDSSVALFEFLNHLPLAGVMSLLATVLMVLF
VTSADSGALVDYLTAKTENSPTWQRLFWTVLMAVLAIIILLVGGLAALQASIIIMSALPF
TFVMLLSWGLLKALKIDVTMKAIQEARITPRAIHNPRSQRLGLIMHYPHTEAEVRE
YIETQVRKAFESIQKEFKRRHLEVNIQNTTEQGLQLRVDHHEINFIYTVYAHETVPPSFM
LGGQDADEAQYFQAEVFLREGGQNYDVMQWTDQEDLIQDIIDQYERHLYFLNIVRGTS
>fig|1310800.6.peg.2 hypothetical protein [Acinetobacter baumannii 1000]
MINNAQSDCDDQEIYFIKSIYINLFIHLLQHYNVIVDIFYTRARVLHCMSSLLFTL
>fig|1310800.6.peg.3 L-carnitine dehydratase/bile acid-inducible protei
1000160 | 1310800.6]
MGALTGLRVLDLSRVLGAPWCSQILADLGAEVIKIERPQTGDDTRMWPWPKDDAGQPT
REAGYYQSANRNKYSVTIDLASKQGQALLREVMKSDVLIENYKVGSLHXYGLDYASLRE
INPGLVYCSITGFGQTGPRASEPGYDFVIQGMAGLMSITGEKDDLPGGGPQKVGAVTDL
QTGLYATIAIQAALWSRQSTGKGQHIDL SLLDVQVAGLANHGMNYLITGQAPQRWGNAHA
NIVPYQTFRAKDKDFIACGNDAQFKALCQAIQAEWAHPKFRNLNQDRIARHRELTTLM
DQHFLTQSAQYWVERIHVVKVPVGMINNVEEALQEPQVQAREMVVSIHPQYRHFKMIAS
PLKLSATPVEYRLAPPRLGEHTQQVLAQYCSTEHLAELTEQGVIGH
>fig|1310800.6.peg.4 Sialic acid transporter (permease) NanT [Acinetoba

```

### Step 6. How do we find our private genome on the PATRIC website?

- When you annotate a private genome in PATRIC genome, it gets indexed in our database, but only you can see it
- The ability to share genomes is coming soon
- To see your private genomes, First click on the "Organisms" tab at the top of the page
- Then click on "All Bacteria"



- This will bring up the overview page for all bacteria in the database
- Click the "Genome List" tab



- This will bring up the list of all genomes
- You can filter the genome list by clicking on the checkboxes on the left hand panel
- Click the checkbox that says "Private"

cellular organisms • Bacteria

Overview Taxonomy Phylogeny **Genome List** Feature Table Specialty Genes Pathways Transcriptomics Interactions Diseases Literature

The Genome List provides all of the available genomic level. To learn more about available metadata, see [Genome](#)

Genome Lists contain a summary list of all genomes associated with a given Phylum, Class, Order, Family, Genus or Species.

Filter By: keyword Clear All

Public (32763)

- Public (32761)
- Private (2)

Genome Status (32763)

- WGS (29276)
- Complete (3481)
- Plasmid (6)

Reference Genome (2601)

32763 genomes found

Workspace Download Columns

Add Genome(s) Table Show/Hide Default

Genome Name	Genome Status	Genome Brows	PATRIC CDS	Isolation Count	Host Name	Disease
'Chrysanthemum coronariu...	WGS		820	Japan	Chrysanthe...	
'Nostoc azollae' 0708	Complete		7014			
Abiotrophia defectiva ATCC...	WGS		3118		Human, Ho...	

- This will filter the list to the two Acinetobacter genomes that we annotated

Filter By: keyword Clear All

Public (32763)

clear

- Public (32761)
- Private (2)

Genome Status (2)

- WGS (2)

Reference Genome (0)

2 genomes found

Workspace Download Columns

Add Genome(s) Table Show/Hide Default

Genome Name	Genome Status	Genome Brows	PATRIC CDS	Isolation Count	Host Name	Disease	Collection Date	Completion Date
Acinetobacter baumannii 10...	WGS		3779					
Acinetobacter baumannii 15...	WGS		3891					

- Typically, when you browse the public genomes, you can select columns
- Being able to add these columns for private genomes is coming soon

## Step 7. Let's look at the genomes

- First click on one of the genomes
- This will load the genome landing page

Bacteria • Proteobacteria • Gammaproteobacteria • Pseudomonadales • Moraxellaceae • Acinetobacter • Acinetobacter baumannii 1000160

Overview Phylogeny Genome Browser Circular Viewer Feature Table Specialty Genes Pathways Protein Families Literature

Search Tools Genome Summary

GF Genome Finder Add Genome to Workspace

FF Feature Finder

CP Comparative Pathway Tool

PFS Protein Family Sorter

Recent PubMed Articles

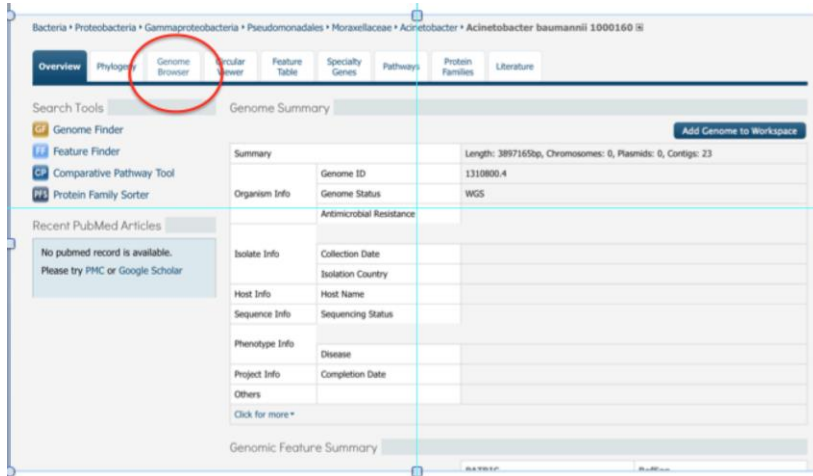
No pubmed record is available. Please try PMC or Google Scholar

Summary	Length: 3897165bp, Chromosomes: 0, Plasmids: 0, Contigs: 23	
Organism Info	Genome ID	1310800.4
	Genome Status	WGS
Isolate Info	Antimicrobial Resistance	
	Collection Date	
	Isolation Country	
Host Info	Host Name	
Sequence Info	Sequencing Status	
Phenotype Info	Disease	
Project Info	Completion Date	
Others		

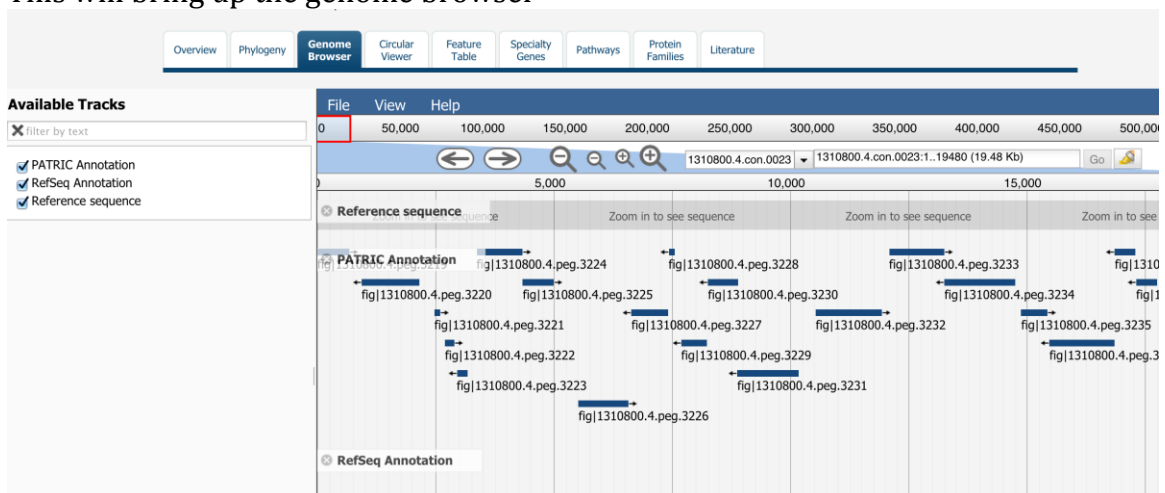
Click for more

Genomic Feature Summary

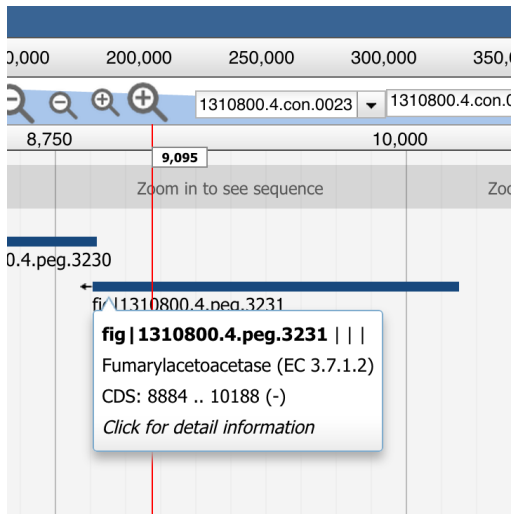
At the top of the page click on the "Genome Browser" tab



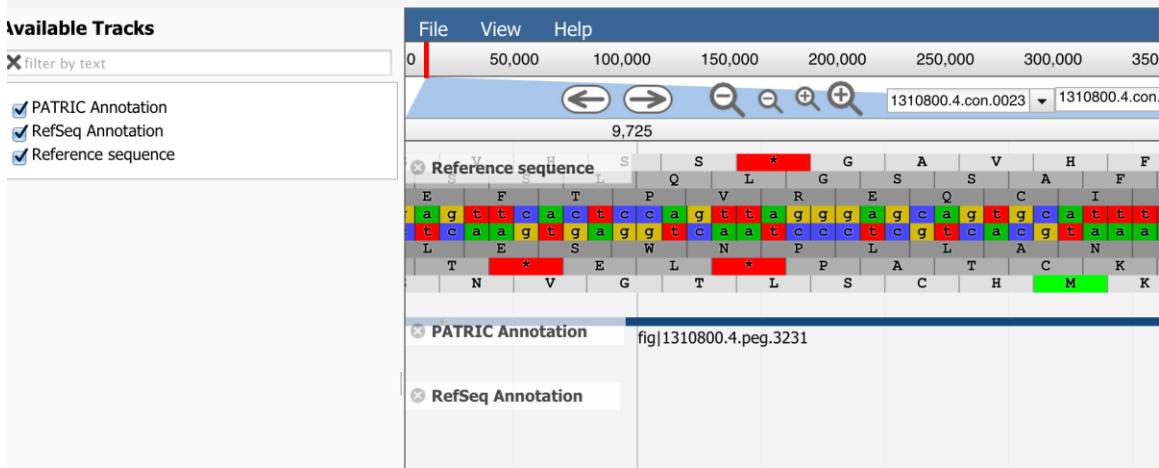
- This will bring up the genome browser



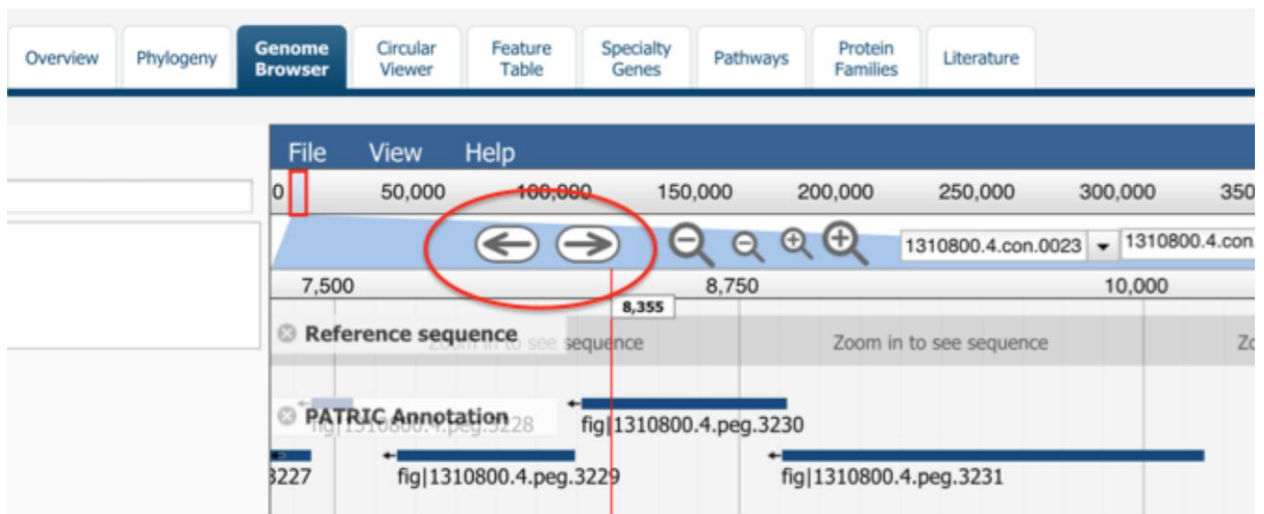
- Doing a mouse over will show the annotation of the gene.
- Arrows display direction of transcription.



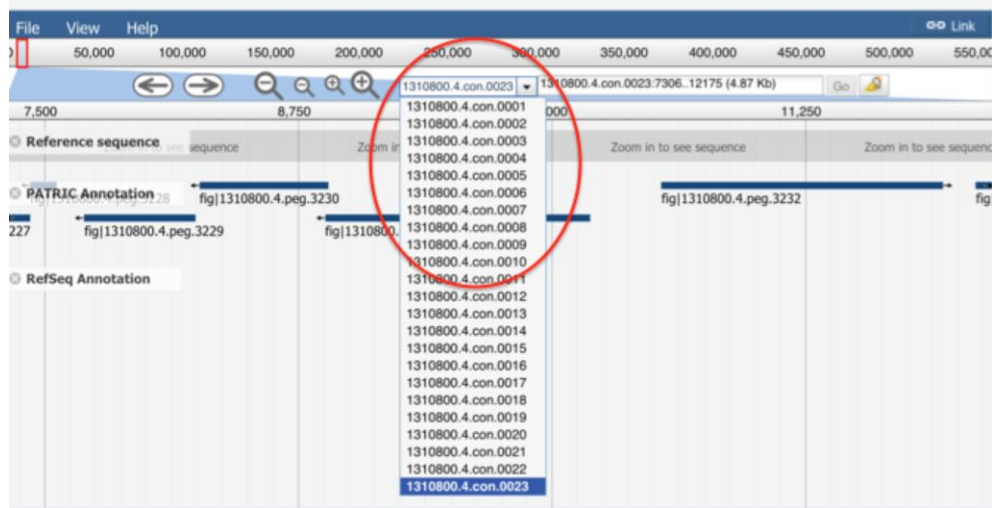
- You can zoom to the sequence by clicking on the magnifying glass icon



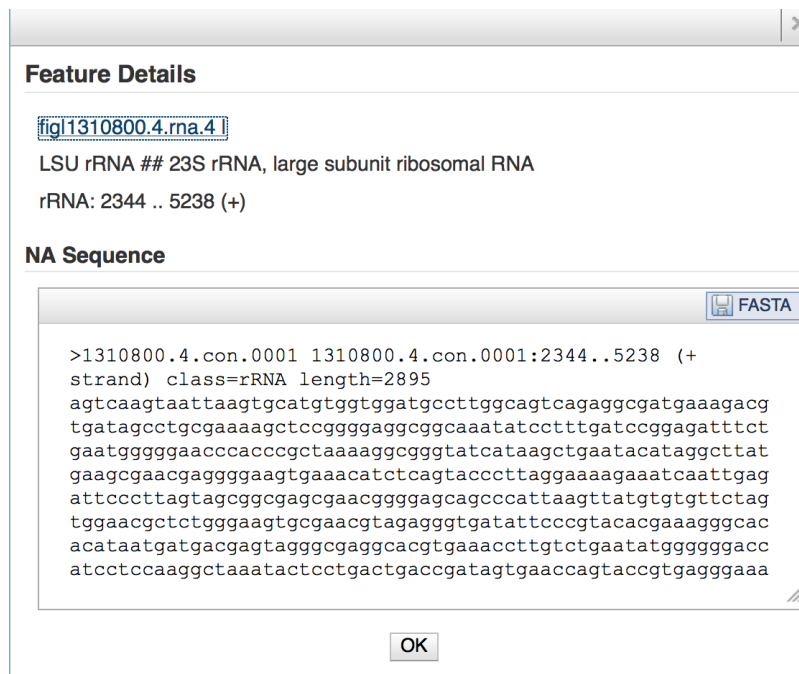
- You can walk the contig by clicking on the arrow buttons



- You can switch contigs using the dropdown menu



- If you click on a gene, it will display the sequence in a pop up



- You can click on the "feature table" tab to view a table of all features in the genome

Feature tables contain all of the identified features for all of the genomes in a particular genus. Tables may be refined to show subsets of features via various user controls, as described in [Feature Table FAQs](#).

Feature Type: CDS Annotation: PATRIC Keyword:

**1779 features found**

PATRIC ID	RefSeq Locus Tag	Alt Locus Tag	Gene Symbol	FIGfam ID	Product Description
<input type="checkbox"/> fig1310800.4.pep.1				FIG00001690	High-affinity choline uptake protein BeT
<input type="checkbox"/> fig1310800.4.pep.2					hypothetical protein
<input type="checkbox"/> fig1310800.4.pep.3				FIG01318850	L-carnitine dehydratase/bile acid-inducible protein F
<input type="checkbox"/> fig1310800.4.pep.4				FIG00004149	Sialic acid transporter (permease) NanT
<input type="checkbox"/> fig1310800.4.pep.5				FIG01332484	Glutaryl-CoA dehydrogenase (EC 1.3.99.7)
<input type="checkbox"/> fig1310800.4.pep.6				FIG01304134	Transcriptional regulator, LysR family
<input type="checkbox"/> fig1310800.4.pep.7				FIG00006661	Potassium channel protein

- You can perform a keyword search to filter the table.

Feature Type: CDS Annotation: PATRIC Keyword:

**82 features found**

PATRIC ID	RefSeq Locus Tag	Alt Locus Tag	Gene Symbol	FIGfam ID	Product Description
<input type="checkbox"/> fig1310800.4.pep.81				FIG00001028	Ribosomal large subunit pseudouridine synthase A (EC 4.2.1.70)
<input type="checkbox"/> fig1310800.4.pep.174				FIG00000471	LSU ribosomal protein L34p
<input type="checkbox"/> fig1310800.4.pep.191				FIG00002492	UDP-N-acetylmuramate-L-alanyl-gamma-D-glutamyl-meso-diamin...
<input type="checkbox"/> fig1310800.4.pep.247				FIG00000213	SSU ribosomal protein S20p
<input type="checkbox"/> fig1310800.4.pep.314				FIG01303917	Similar to ribosomal large subunit pseudouridine synthase A
<input type="checkbox"/> fig1310800.4.pep.317				FIG00002716	Ribosomal small subunit pseudouridine synthase A (EC 4.2.1.70)

- Let's look at those weird specialty genes that we annotated
- Click on the "Specialty Genes" tab

For this taxonomy/genome level, the list below provides Specialty Genes, i.e. genes that are of particular interest to the infectious disease researchers, such as virulence factors, antibiotic resistance genes, drug targets, and human homologs. **Genes with the designation "Literature" in the Evidence field have been experimentally verified. Those with Evidence designated as "BLASTP" are identified based on sequence homology.** For more details, see [Specialty Genes FAQs](#).

**Specialty Genes**

102 records found

Evidence	Property	Source	PATRIC ID	RefSeq Loc	Alt Locus Tag	Source ID	Gene	Product	Pubmed	Identity	E-value
<input type="checkbox"/> BLASTP	Drug Target	DrugBank	fig1310800.4.pep.5			6884	Glutaryl-CoA dehydr...			80	0
<input type="checkbox"/> BLASTP	Human Homolog	Human	fig1310800.4.pep.5			NP_000...	Glutaryl-CoA dehydr...			63	0
<input type="checkbox"/> BLASTP	Human Homolog	Human	fig1310800.4.pep.31			NP_001...	ATP synthase beta c...			69	0
<input type="checkbox"/> BLASTP	Human Homolog	Human	fig1310800.4.pep.33			NP_001...	ATP synthase alpha ...			57	0
<input type="checkbox"/> BLASTP	Human Homolog	Human	fig1310800.4.pep.68			NP_004...	Enoyl-CoA hydratase...			55	0
<input type="checkbox"/> BLASTP	Human Homolog	Human	fig1310800.4.pep.72			NP_005...	Methylmalonate-se...			60	0.0
<input type="checkbox"/> BLASTP	Human Homolog	Human	fig1310800.4.pep.1...			NP_000...	UDP-glucose 4-epim...			54	2e-123
<input type="checkbox"/> BLASTP	Human Homolog	Human	fig1310800.4.pep.1...			NP_076...	LSU ribosomal prote...			53	0.0000...
<input type="checkbox"/> BLASTP	Human Homolog	Human	fig1310800.4.pep.1...			NP_079...	Acyl-CoA dehydroge...			51	2e-133

- Membership in this list is based on BLAST similarity to a reference database
- You can filter the list by database, blast score, and closest matching organism



The screenshot shows a bioinformatics tool interface. On the left, there are search filters under 'Properties (102)', 'Source (27)', and 'Evidence (27)'. The 'Antibiotic Resistance (27)' filter is selected. Below these are 'Filter BLAST Hits by:' options for 'Source organism' (Same Genome, Same Species, Same Genus, All) and percentage thresholds for 'Query Coverage', 'Subject Coverage', and 'Identity'. The main table displays search results with columns: Evidence, Property, Source, PATRIC ID, RefSeq Loc, Alt Locus Tag, Source ID, Gene, Product, Pubmed, Identity, and E-value. The table lists 27 results, all related to Antibiotic Resistance from the CARD source.

- If you click on the "Pathways" tab you can see all the genes with annotations that map to KEGG

The screenshot shows the 'Pathways' tab selected in a bioinformatics tool. The interface includes a navigation bar with tabs: Overview, Phylogeny, Genome Browser, Circular Viewer, Feature Table, **Pathways**, Protein Facts, and Literature. Below the navigation bar, there are search filters for 'Pathway Class', 'Pathway Name', 'EC Number', and 'Annotation'. The main content area displays '137 unique pathway(s) found' and a table with columns: Pathway ID, Pathway Name, Pathway Class, Annotation, Unique Genome Count, Unique Gene Count, Unique EC Count, EC Conservation, and Gene Conservation. The table lists various metabolic pathways such as Glycolysis / Gluconeogenesis, Citrate cycle (TCA cycle), and Fatty acid biosynthesis.

- If you click on one of the pathways it will bring up the KEGG map

Pathways EC Numbers Genes

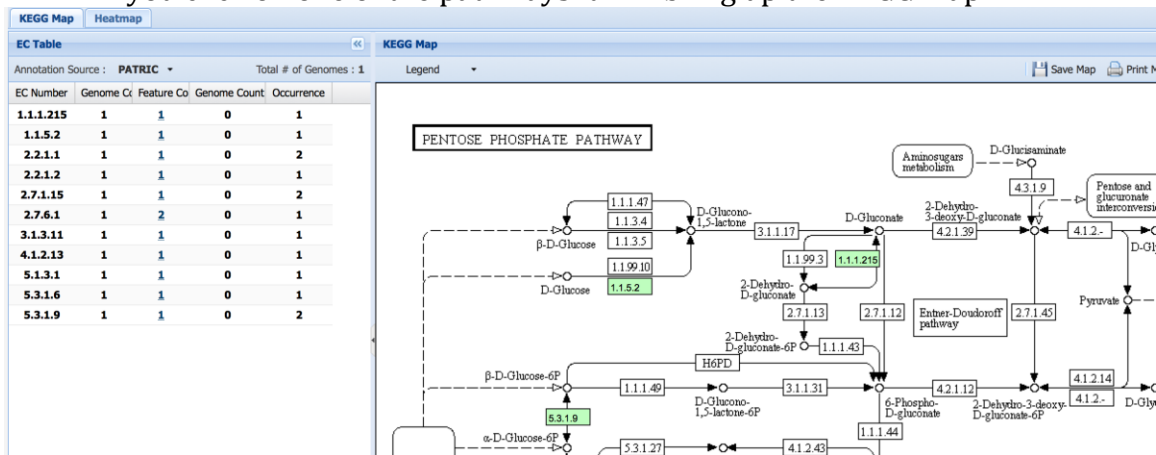
**137 unique pathway(s) found**

Workspace View Download Tools

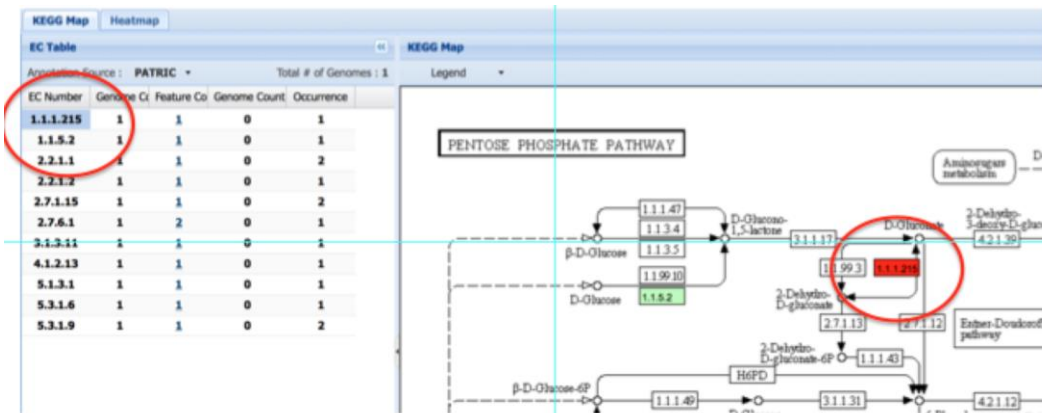
Add Feature(s) FASTA DNA FASTA Protein Table FASTA Pathway Summary Multiple Se MAP IDs to...

Pathway ID	Pathway Name	Pathway Class
00010	Glycolysis / Gluconeogenesis	Carbohydrate Metabolism
00020	Citrate cycle (TCA cycle)	Carbohydrate Metabolism
00030	Pentose phosphate pathway	Carbohydrate Metabolism
00040	Pentose and glucuronate interconversions	Carbohydrate Metabolism
00051	Fructose and mannose metabolism	Carbohydrate Metabolism
00052	Galactose metabolism	Carbohydrate Metabolism

- If you click on one of the pathways it will bring up the KEGG map



- If you click an EC number in the side bar it will highlight the enzyme in the map



- Finally you can render a diagram of the genome.
- Click on the "Circular Viewer" tab.



- You can control the layout by clicking the checkboxes.
- You can download a publication-quality image by clicking the download button.

