

Using the Proteome Comparison Tool in PATRIC

In the last five years, many studies have focused on pan genome comparisons between bacterial strains. In order to perform one of these comparisons the researcher must know which genes are "core" or shared between strains, and which genes are "variable" or idiosyncratic to a given strain. In order to enable these types of analyses, we have deployed the proteome comparison tool from RAST as a service in PATRIC. The tool allows you to compute bidirectional best hits between a reference genome and up to 8 other genomes at a time. In its current implementation, the tool runs a BLASTP comparison, but in future releases we intend to enable BLASTN comparisons so that users can compare DNA features like RNA genes, and multi-directional best hit analyses.

In this example, we will compare the two *A. baumannii* strains that we annotated by when we completed the annotation tutorial to the *A. baumannii* AYE strain that is publicly available in PATRIC. AYE is particularly interesting because it caused an outbreak of nosocomial infections in Europe.

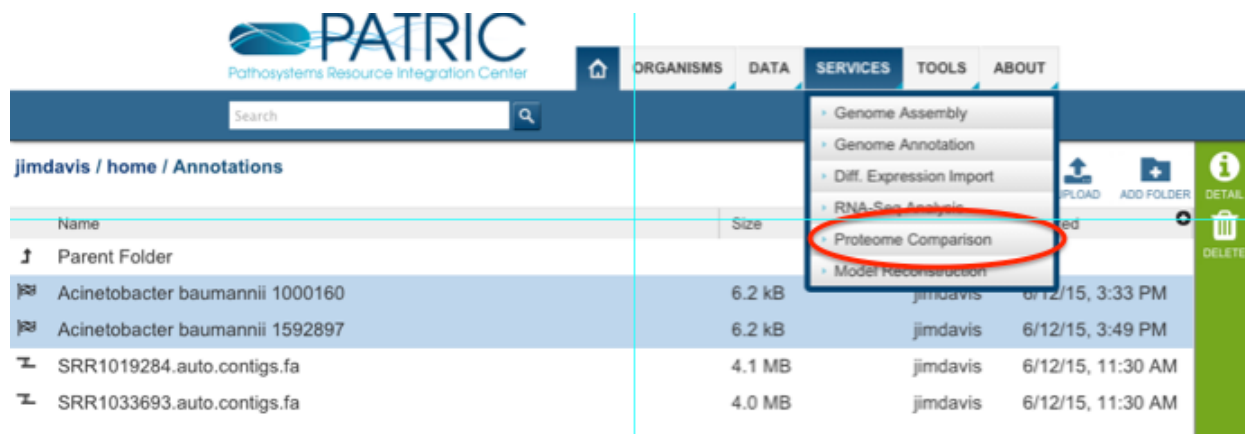
First, log into your workspace. You should have a directory of annotated genomes for *A. baumannii* 1000160 and *A. baumannii* 1592897.

jimdavis / home / Annotations

Name	Size	Owner	Created
Parent Folder			
Acinetobacter baumannii 1000160	6.2 kB	jimdavis	6/12/15, 3:33 PM
Acinetobacter baumannii 1592897	6.2 kB	jimdavis	6/12/15, 3:49 PM
SRR1019284.auto.contigs.fa	4.1 MB	jimdavis	6/12/15, 11:30 AM
SRR1033693.auto.contigs.fa	4.0 MB	jimdavis	6/12/15, 11:30 AM

If you do not have these annotated genomes, you can find them in the shared workshopdemo workspace.

Next, select the proteome comparison tool from the Services tab.



This will bring up the proteome comparison app.

Proteome Comparison
Protein sequence-based comparison using bi-directional BLASTP.

Parameters ⓘ

Reference Genome
▼ e.g. *Mycobacterium tuberculosis* H37Rv

Add Up To 8 Genomes to Compare (use plus button to add)
▼ e.g. *M. tuberculosis* CDC1551 +

Optional External Genome (Protein Fasta file)
Optional

Advanced ▾

OUTPUT FOLDER

OUTPUT NAME
Output Name

Reset Submit

The app is designed to compare up to 8 genomes to a reference genome. We will select *Acinetobacter baumannii* AYE as our reference strain by typing it into the first field and allowing the list to autocomplete.

Parameters ⓘ

Reference Genome
▼ Acinetobacter baumannii AY
Acinetobacter baumannii AYE VEB [509173.8]
Acinetobacter baumannii UH8907 [1398988.3]
Acinetobacter baumannii UH6507 [1398977.3]
Acinetobacter baumannii UH5107 [1398971.3]
Acinetobacter baumannii UH0207 [1398943.3]
Acinetobacter baumannii UH0807 [1398945.3]
Acinetobacter baumannii UH0707 [1398944.3]
Acinetobacter baumannii UH9907 [1398991.3]
Acinetobacter baumannii UH9007 [1398989.3]
Acinetobacter baumannii UH8107 [1398984.3]

Optional External Genome (Protein Fasta file)
Optional

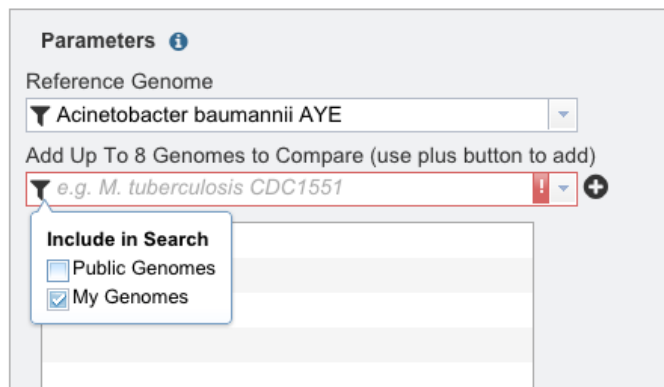
Advanced ▾

OUTPUT FOLDER

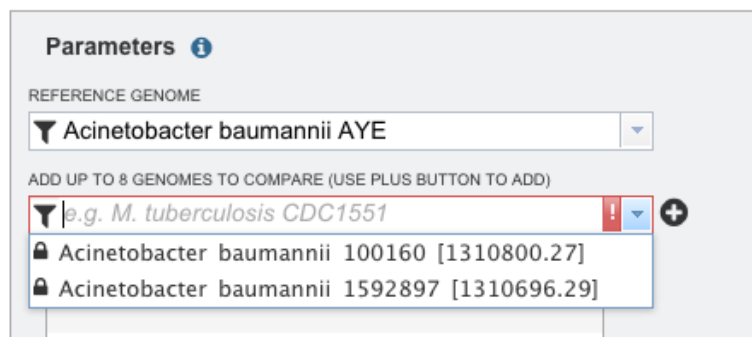
OUTPUT NAME
Output Name

Genome

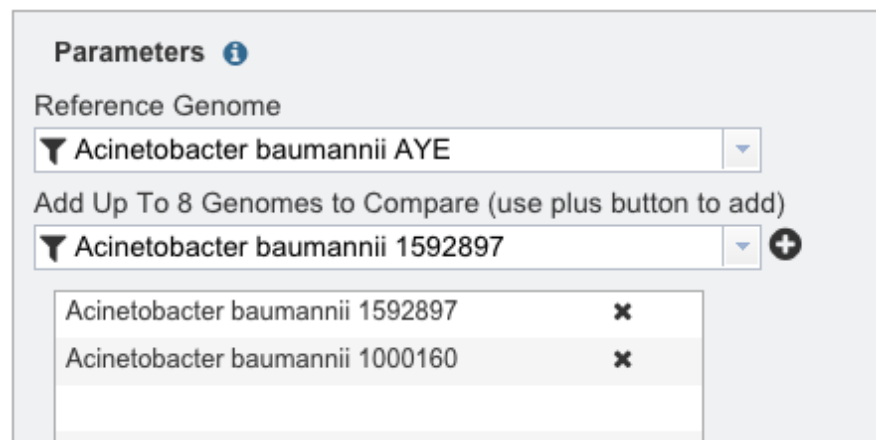
Then we will select the two strains that we just sequenced. It's easiest to find these by limiting the search to private genomes by clicking on the filter icon in the search box and deselecting the public genomes.



Then we will only see the private genomes from the workspace.



Add 1592897 and 1000160 to the comparison by clicking the plus sign to the right of the dropdown box.



This tool also allows you to specify an external genome if you wish to add one which is not in your workspace. The genome must be in fasta format. If you click the arrow button to expand the advanced options, it exposes the BLAST parameters. For this tutorial we will

use the default settings. We will name the output folder, "Proteome_Comparison", and the output file as, "AYE_vs_159287_and_1000160". Then we will click submit.

Proteome Comparison
Protein sequence-based comparison using bi-directional BLASTP.

Parameters ⓘ

Reference Genome

Add Up To 8 Genomes to Compare (use plus button to add)
 +

Acinetobacter baumannii 1592897 ✕

Acinetobacter baumannii 1000160 ✕

Optional External Genome (Protein Fasta file)

Advanced ⌵

MINIMUM % COVERAGE: BLAST E-VALUE:

OUTPUT FOLDER

OUTPUT NAME



Just like with the other apps, the job will show up on the jobs page.

Status	Submit	App	Output Name	Start	Completed
● in-progress	7/17/15, 2:58 PM	GenomeComparison	AYE_vs_159287_and_1000160	7/17/15, 2:58 PM	
● completed	7/14/15, 8:58 AM	GenomeComparison	1000160_vs_SDF_and_AYE	7/14/15, 8:58 AM	7/14/15, 9:04 AM
● completed	7/14/15, 8:14 AM	ModelReconstruction	A_baumannii_gmm_1	7/14/15, 8:17 AM	7/14/15, 8:19 AM
● completed	7/14/15, 8:14 AM	ModelReconstruction	A_baumannii_complete_1	7/14/15, 8:14 AM	7/14/15, 8:16 AM

There are two ways to get to the output when the job has completed. First you can double click on the row in the jobs page.

Status	Submit	App	Output Name	Start
● completed	7/17/15, 2:58 PM	GenomeComparison	AYE_vs_159287_and_1000160	7/17/15, 2:58 PM

Second you can find the "Proteome_Comparison" directory in your workspace and click on the icon for the job.

Name	Size	Owner	Created
↑ Parent Folder			
 1000160_vs_SDF_and_AYE	4.9 kB	jimdavis	7/14/15, 9:04 AM
 AYE_vs_159287_and_1000160	4.9 kB	jimdavis	7/17/15, 3:01 PM

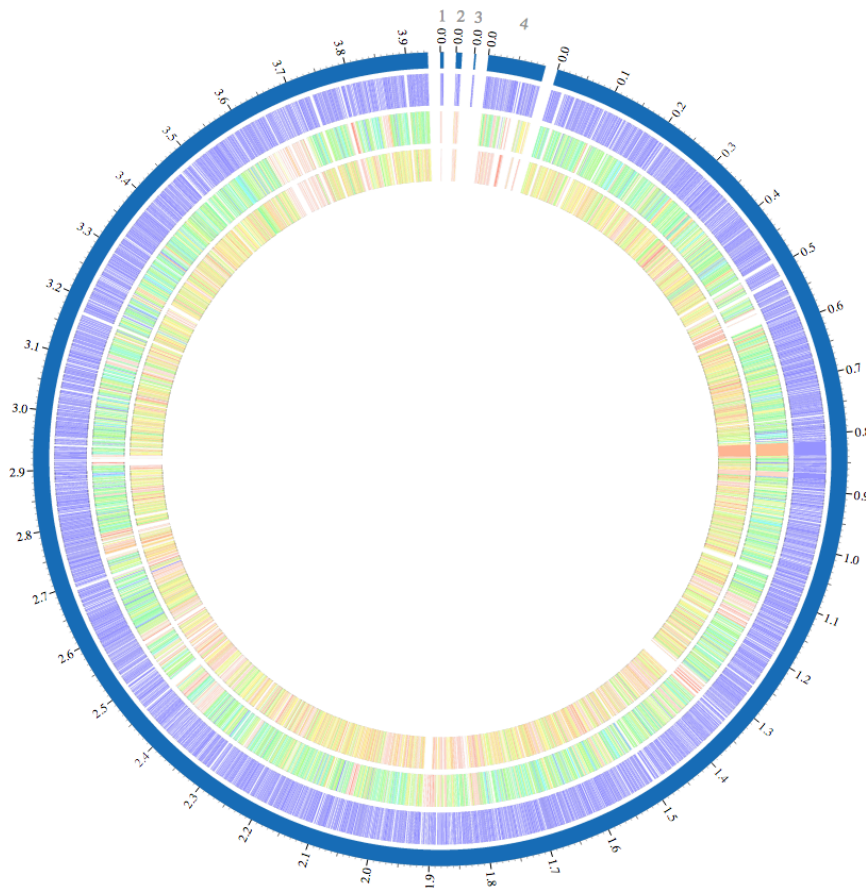
This will load the proteome comparison output.

The comparison is displayed by the circos tool and proteins are aligned against the AYE (reference strain) contigs.

	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10
Bidirectional best hit	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10
Unidirectional best hit	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10

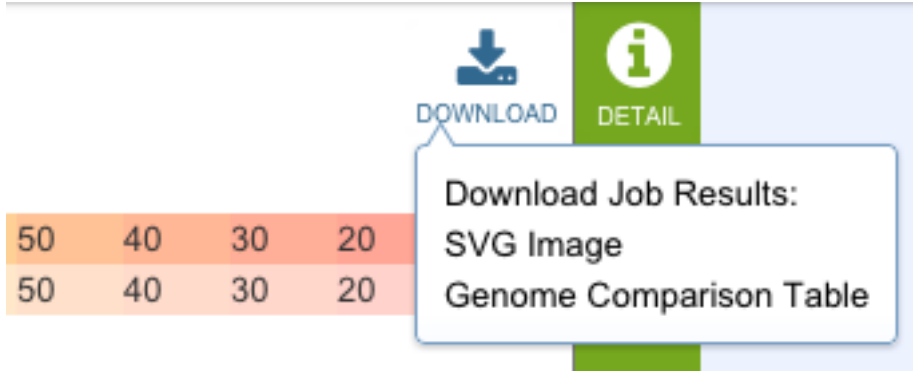
List of tracks, from outside to inside:

1. Acinetobacter baumannii AYE (509173.8)
2. Acinetobacter baumannii 1592897 (1310696.15)
3. Acinetobacter baumannii 1000160 (1310800.18)



Track order is shown in the top left, and the AYE strain is the outside track. Using the color key you can get a feeling for the similarity of each protein over the chromosome. It looks like the AYE strain is more similar to the 1592897 strain than it is to the 1000160 strain

because its track has more proteins that are greens and blues, and the 1000160 has more yellows and oranges. You can download the circular diagram as a high resolution SVG graphic if you are preparing a manuscript. You can also study the output in more detail by downloading the Genome Comparison Table.



The genome comparison table is a tab-delimited flat text file and can be opened in any text editor or imported into excel. The table contains the gene ids, annotations, information on whether they are bidirectional or unidirectional best hits, and BLAST data.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Acinetobacter baumannii AYE										Acinetobacter baumannii 1592897					
ref_genome_contig	ref_genome_gc	ref_genome_aa_length	ref_genome_patric_id	ref_genome	ref_genome	ref_genome	ref_genome	ref_genome	ref_genome	comp_genor	comp_genor	comp_genor	comp_genor	comp_genor	comp_genor
NC_010401	1	316	fig 509173.8.peg.1	p1ABAYE0001	Plasmid repli	629	1579	+	uni	1310696.15.	2814	307	fig 1310696.15.peg.2814		
NC_010401	2	95	fig 509173.8.peg.2	p1ABAYE0002	hypothetical	1576	1863	+							
NC_010401	3	172	fig 509173.8.peg.3	p1ABAYE0003	FIG00354122	2054	2572	-	uni	1310696.15.	1194	296	fig 1310696.15.peg.1194		
NC_010401	4	87	fig 509173.8.peg.4	p1ABAYE0004	YefM protei	2851	3114	+							
NC_010401	5	100	fig 509173.8.peg.5	p1ABAYE0005	hypothetical	3426	3728	-							
NC_010401	6	581	fig 509173.8.peg.6	p1ABAYE0007	hypothetical	3846	5591	+							
NC_010402	7	316	fig 509173.8.peg.7	p2ABAYE0001	replication ir	662	1612	+	bi	1310696.15.	2215	307	fig 1310696.15.peg.2215		
NC_010402	8	191	fig 509173.8.peg.8	p2ABAYE0002	putative prof	1605	2180	+	bi	1310696.15.	2214	181	fig 1310696.15.peg.2214		
NC_010402	9	356	fig 509173.8.peg.9	p2ABAYE0004	Threonine de	2822	3892	-	uni	1310696.15.	1187	387	fig 1310696.15.peg.1187		

You should take some time to traverse the table. You will find a large phage element that is found in AYE and 1000160, but not 1592897 (fig|509173.8.peg.612 to fig|509173.8.peg.647). You will also find a CRISPR-associated region that is unique to AYE (fig|509173.8.peg.1042 to fig|509173.8.peg.1052).