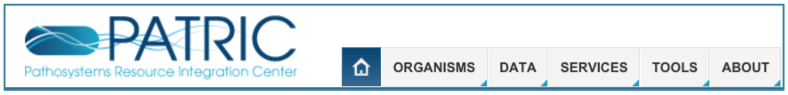
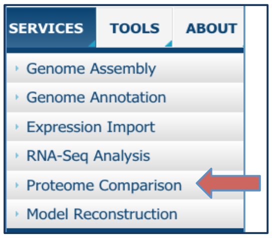
Protein Family Analysis: Proteome Comparison

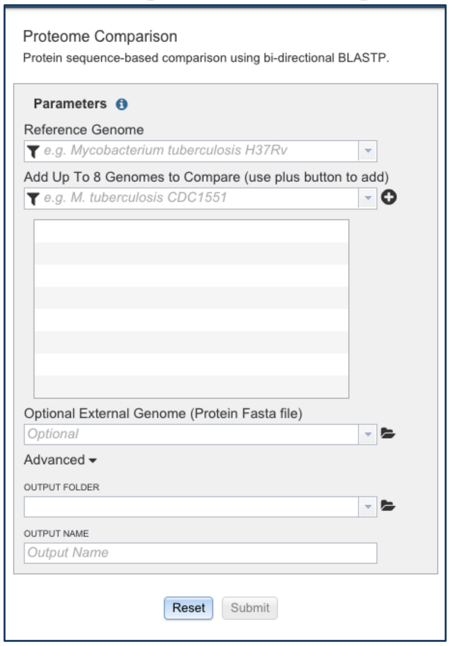
1. At the top of any PATRIC page, find the Services tab and click on it.



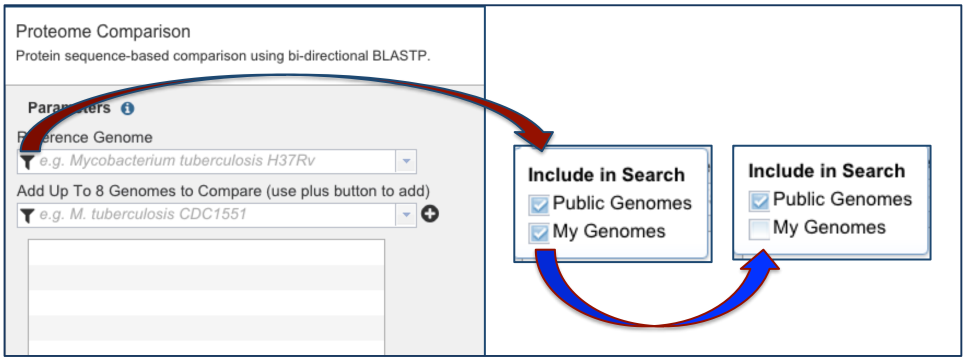
2. Click on Proteome Comparison



3. This will open up the landing page for where you can submit a Proteome Comparison job. This tool is a best bidirectional BLAST comparison of the annotated proteins from up to nine different genomes.



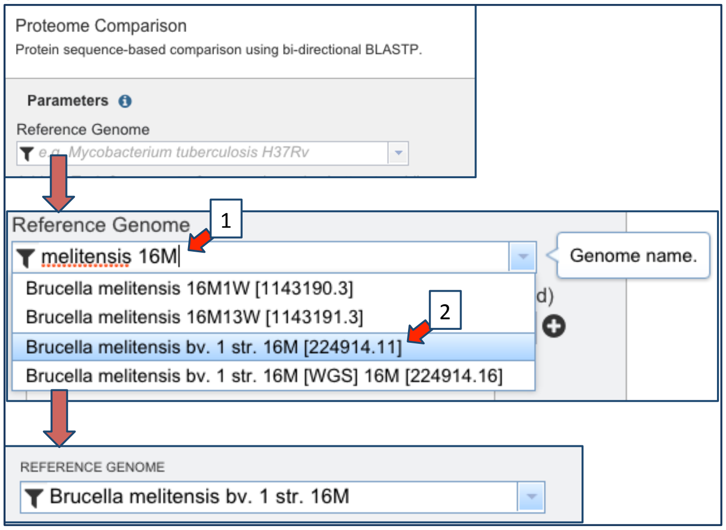
4. **Selecting the Reference Genome.** You will first need to select a genome to serve as the reference for the comparison. Under Reference Genome, click on the filter icon (Red arrow). This will open a box that allows you to search across all of the public genomes available in PATRIC, or across the genomes that you have annotated and that are stored in your workspace. In this example, we will use a public genome as the reference, so click on the box in front of My Genomes (Blue arrow) to deselect that box.



Include in Search
☑ Public Genomes
☑ My Genomes

Include in Search
☑ Public Genomes
☐ My Genomes

5. In the text box below Reference Genome, start typing some words that will identify the reference. You can see that I used "melitensis" and "16M" (Red arrow 1). Once I started typing, a list appeared below the box, providing me with possible choices that match my text. You can see that there are a number of 16M genomes. Once I selected a particular genome by clicking on it (Red arrow 2), the name of that genome auto filled into the text box (lowest panel).

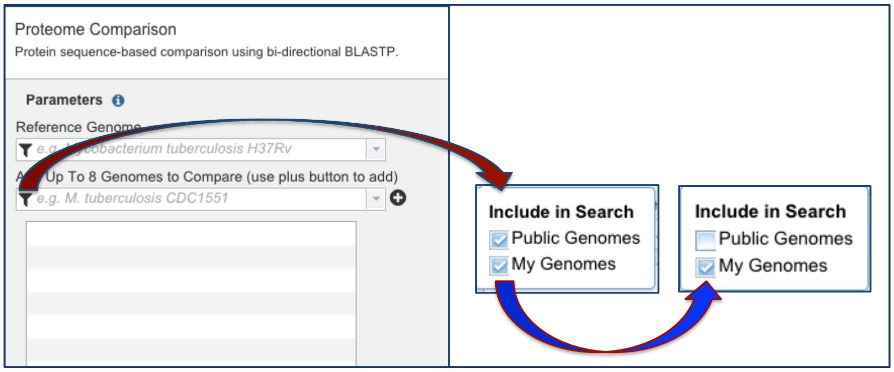


Reference Genome: melitensis 16M

Brucella melitensis 16M1W [1143190.3]
Brucella melitensis 16M13W [1143191.3]
Brucella melitensis bv. 1 str. 16M [224914.11]
Brucella melitensis bv. 1 str. 16M [WGS] 16M [224914.16]

REFERENCE GENOME: Brucella melitensis bv. 1 str. 16M

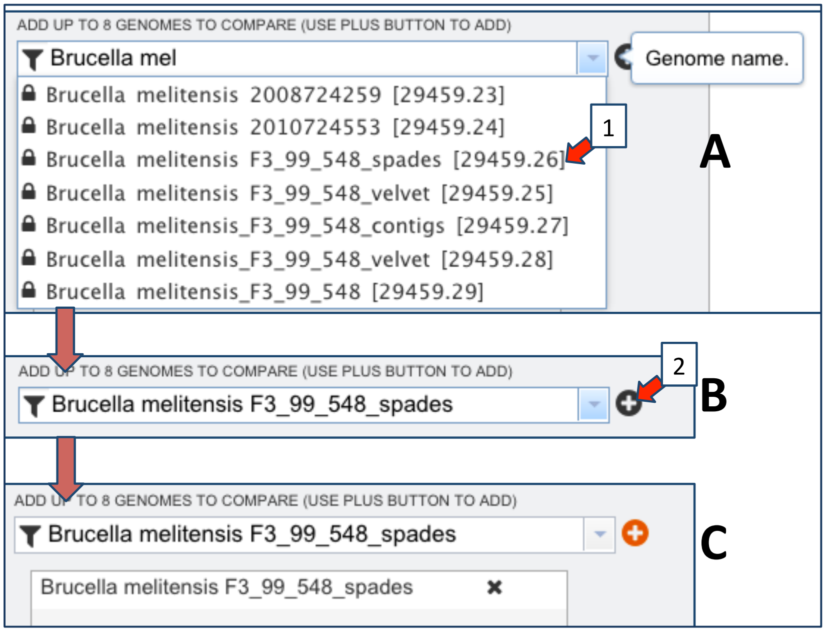
6. **Selecting the Comparison Genomes.** Now you need to select genomes to compare to your reference. You'll need to follow the same steps as in Step 4 above. In this example, I want to start with one of my private genomes, so I deselect the check box in front of Public Genomes.



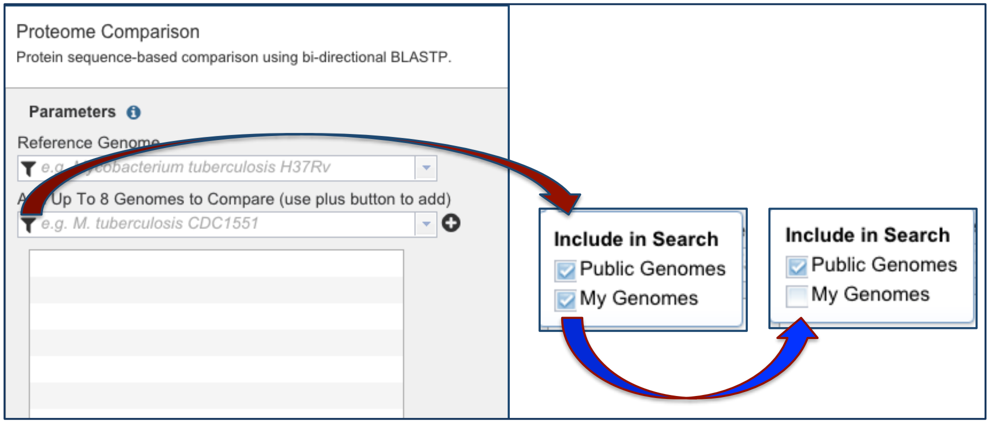
Include in Search
☑ Public Genomes
☑ My Genomes

Include in Search
☐ Public Genomes
☑ My Genomes

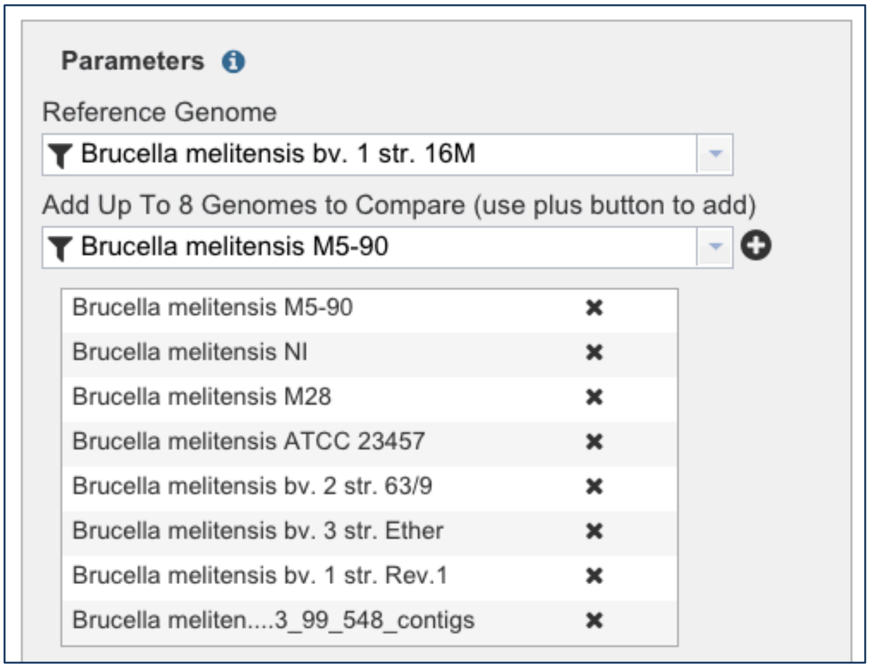
7. Once again, you will need to start typing a name into the text box. Once you have entered enough text to see the genome you are interested in click on that (Red Arrow 1 in Panel A below). This will auto-fill the text box with the name you have selected (Panel B). It that is the right choice, you will need to click on the "+" icon (Red arrow 2). Once you do that, the genome you selected will appear in the box below (Panel C).



8. In this example, we will add in some public genomes for the comparison, to do this you will need to change the filter from private to public genomes. Follow the instructions as per Step 4.



9. Repeat the process from Step 7 until you have made all your selections. Remember to click on the "+" icon after you have chosen each name. All your selections will appear in the large box as seen below.



Parameters

Reference Genome

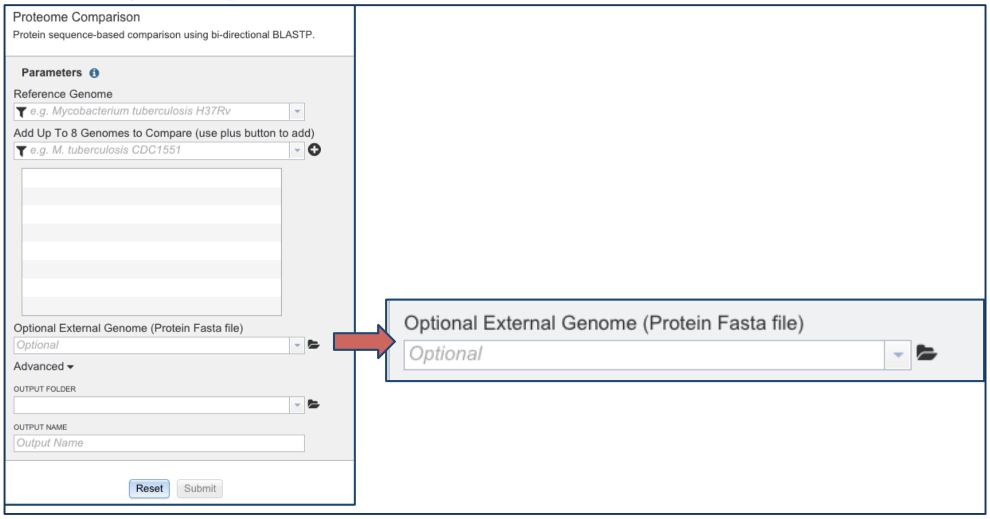
Brucella melitensis bv. 1 str. 16M

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

Brucella melitensis M5-90

- Brucella melitensis M5-90 ✖
- Brucella melitensis NI ✖
- Brucella melitensis M28 ✖
- Brucella melitensis ATCC 23457 ✖
- Brucella melitensis bv. 2 str. 63/9 ✖
- Brucella melitensis bv. 3 str. Ether ✖
- Brucella melitensis bv. 1 str. Rev.1 ✖
- Brucella meliten....3_99_548_contigs ✖

10. **Uploading an outside annotation.** The Proteome Comparison tool also allows you to compare a genome that has not been annotated in PATRIC. To do this, you would need to click on the folder icon next to Optional External Genome (Protein Fasta File) and upload that file.



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

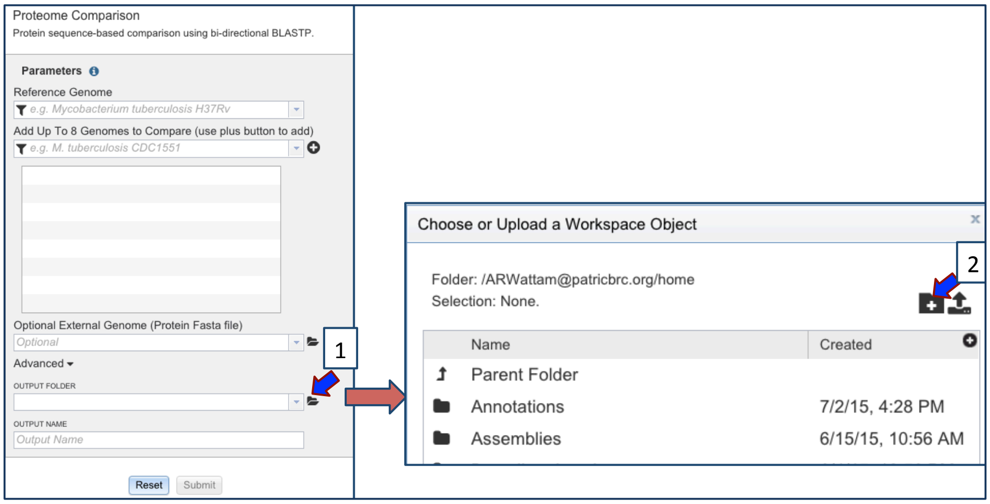
OUTPUT NAME

Output Name

Optional External Genome (Protein Fasta file)

Optional

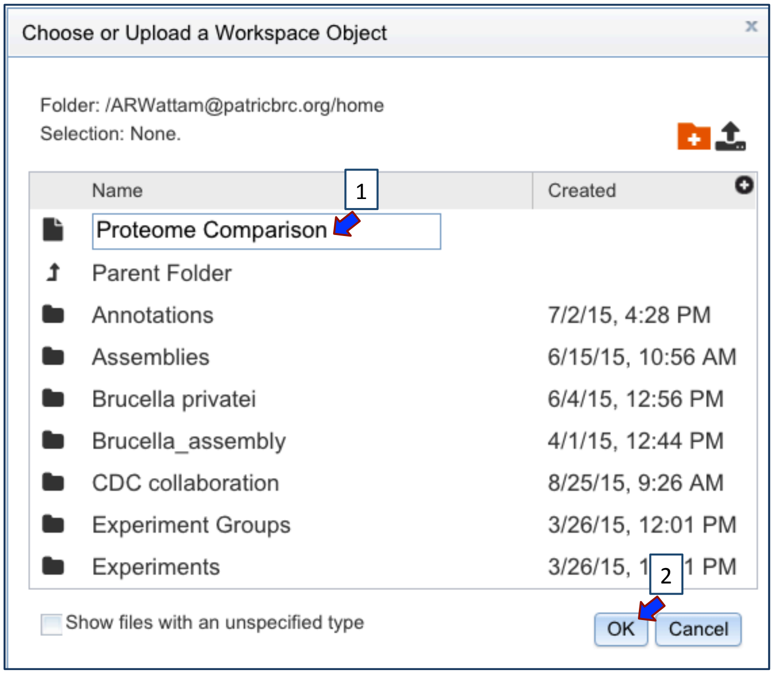
11. **Creating an Output Folder.** Next you will need to choose a folder that will hold the finished comparison. If you haven't done this before, you will need to create a folder. To do this, click on the folder icon that follows Output Folder (Blue arrow 1). This will open up a new window that will show all the folders you have in your workspace. To create a new folder, click on the folder icon with the + sign (Blue arrow 2) that you can see at the top.



Folder: /ARWattam@patricbrc.org/home
Selection: None.

Name	Created
Parent Folder	
Annotations	7/2/15, 4:28 PM
Assemblies	6/15/15, 10:56 AM

12. This will open up a text box that allows you to name the folder you are creating (Blue arrow 1). Once you have entered the name you want, click on the OK button at the bottom of the page (Blue arrow 2). This window will close once you do that.

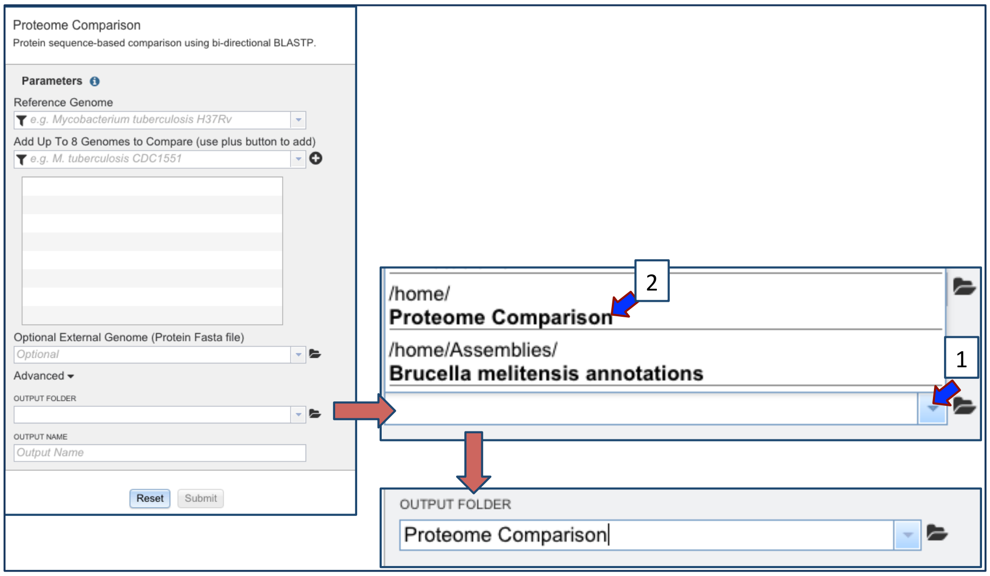


Folder: /ARWattam@patricbrc.org/home
Selection: None.

Name	Created
Proteome Comparison	
Parent Folder	
Annotations	7/2/15, 4:28 PM
Assemblies	6/15/15, 10:56 AM
Brucella privatei	6/4/15, 12:56 PM
Brucella_assembly	4/1/15, 12:44 PM
CDC collaboration	8/25/15, 9:26 AM
Experiment Groups	3/26/15, 12:01 PM
Experiments	3/26/15, 1 1 PM

☐ Show files with an unspecified type

13. Now you need to choose the folder you just created. Click on the down arrow that follows Output Folder (Blue arrow 1). This will open up a new window that will show all the folders you have in your workspace. This will open a drop down box (or drop up in this case) that shows all the possible folders. Select the folder that you want to hold the comparison (Blue arrow 2). This will auto fill the text box under Output Folder with that name.

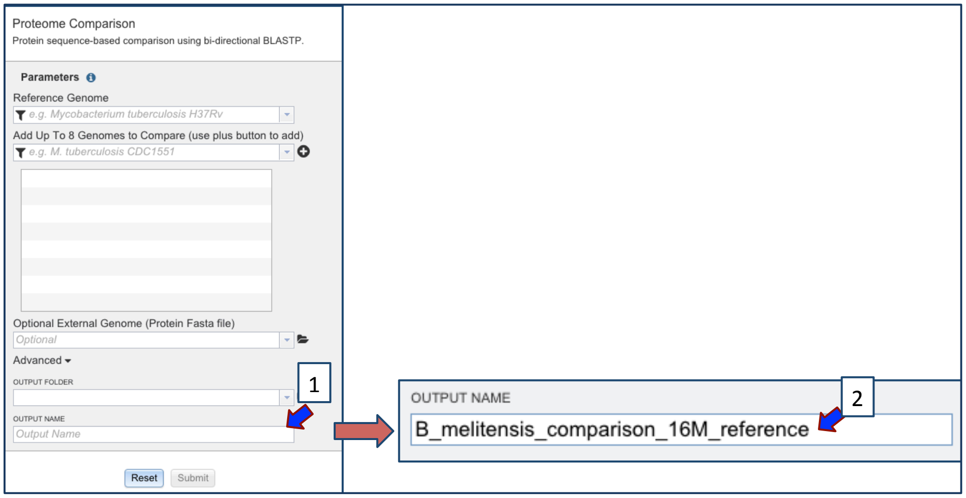


/home/
Proteome Comparison

/home/Assemblies/
Brucella melitensis annotations

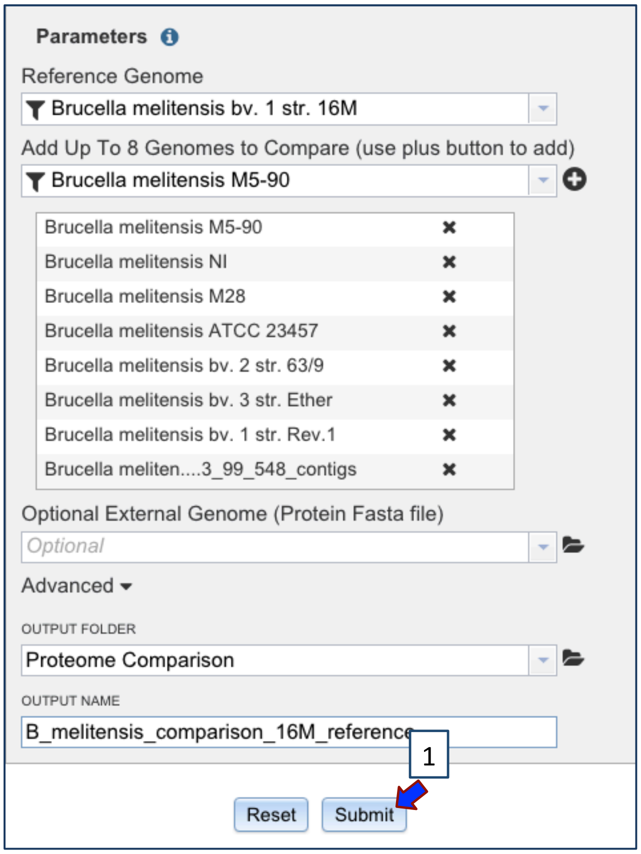
OUTPUT FOLDER
Proteome Comparison

14. Now you need to name your comparison. In the text box under Submit Name (Blue arrow 1) enter a distinctive name that you will remember (Blue arrow 2). Trust me, I do a lot of these and you want a name that describes the comparison.



OUTPUT NAME
B_melitensis_comparison_16M_reference

15. Now you can click the Submit button at the bottom of the page (Blue arrow 1).



Parameters

Reference Genome: Brucella melitensis bv. 1 str. 16M

Add Up To 8 Genomes to Compare (use plus button to add): Brucella melitensis M5-90

- Brucella melitensis M5-90 ✖
- Brucella melitensis NI ✖
- Brucella melitensis M28 ✖
- Brucella melitensis ATCC 23457 ✖
- Brucella melitensis bv. 2 str. 63/9 ✖
- Brucella melitensis bv. 3 str. Ether ✖
- Brucella melitensis bv. 1 str. Rev.1 ✖
- Brucella meliten....3_99_548_contigs ✖

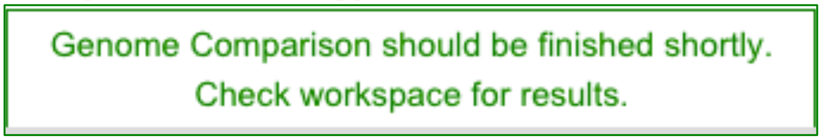
Optional External Genome (Protein Fasta file): Optional

Advanced ▾

OUTPUT FOLDER: Proteome Comparison

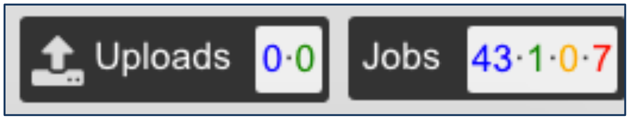
OUTPUT NAME: B_melitensis_comparison_16M_reference

16. A message will appear that tells you your job has been submitted. This message is temporal and will disappear after several seconds.

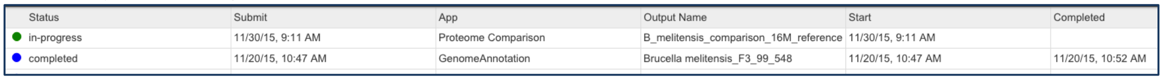


Genome Comparison should be finished shortly.
Check workspace for results.

17. You can check the status of your annotation job by clicking on the Jobs indicator at the bottom left of the PATRIC page.

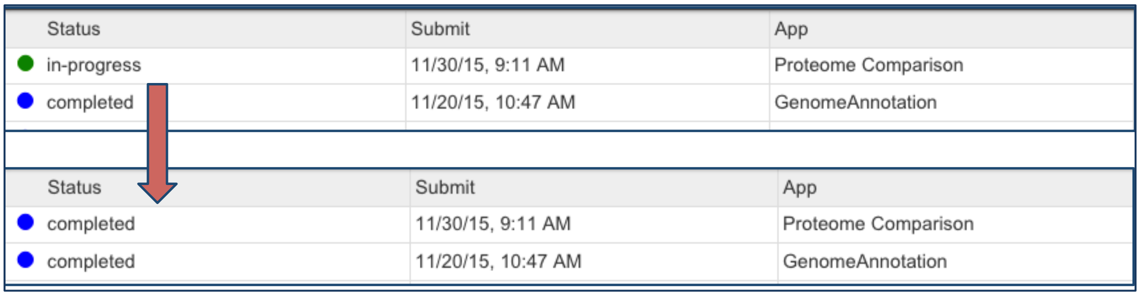


18. Clicking on Jobs will take you to the Jobs Status page, where you can see how your assembly job is progressing. You can also see the status of all the previous service jobs that you have submitted to PATRIC.



Status	Submit	App	Output Name	Start	Completed
● in-progress	11/30/15, 9:11 AM	Proteome Comparison	B_melitensis_comparison_16M_reference	11/30/15, 9:11 AM	
● completed	11/20/15, 10:47 AM	GenomeAnnotation	Brucella melitensis_F3_99_548	11/20/15, 10:47 AM	11/20/15, 10:52 AM

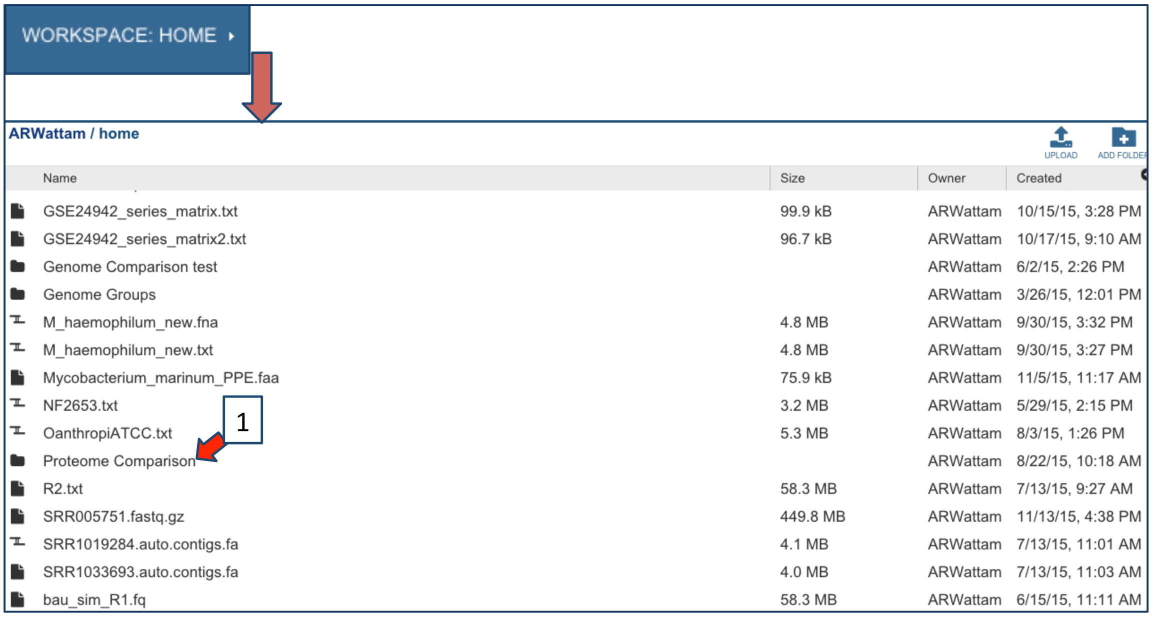
19. You will be able to see when your job is complete.



	Status	Submit	App
●	in-progress	11/30/15, 9:11 AM	Proteome Comparison
●	completed	11/20/15, 10:47 AM	GenomeAnnotation

	Status	Submit	App
●	completed	11/30/15, 9:11 AM	Proteome Comparison
●	completed	11/20/15, 10:47 AM	GenomeAnnotation

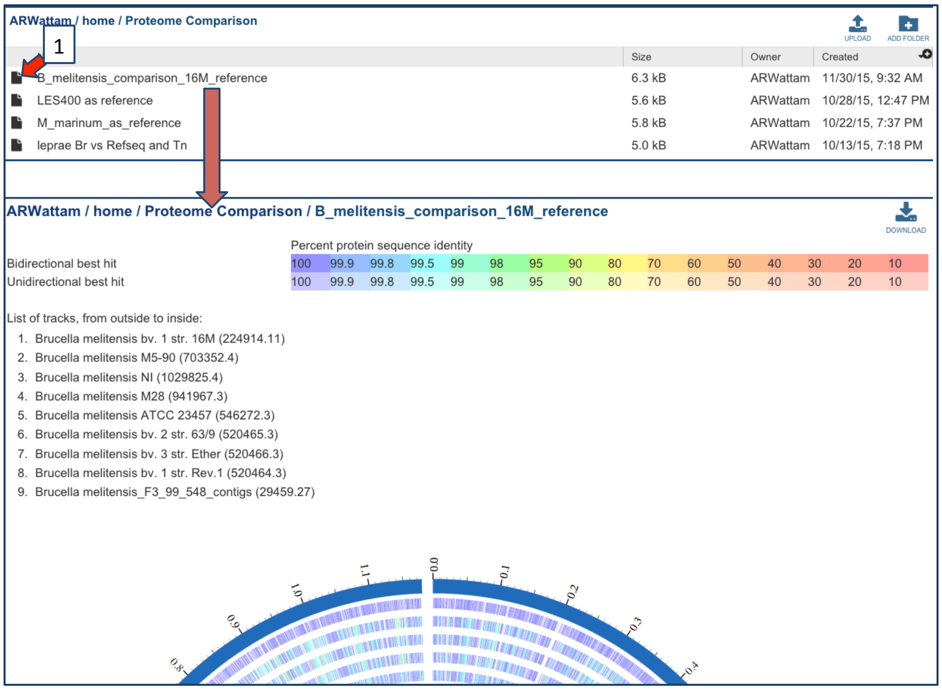
20. To find your job, you will need to click on Workspace Home that you can find at the top left of any PATRIC page. This will open up your workspace, where you can see all your folders. Find the folder where you placed your job and click on the folder icon that precedes the name (Red arrow 1).



ARWattam / home

Name	Size	Owner	Created
GSE24942_series_matrix.txt	99.9 kB	ARWattam	10/15/15, 3:28 PM
GSE24942_series_matrix2.txt	96.7 kB	ARWattam	10/17/15, 9:10 AM
Genome Comparison test		ARWattam	6/2/15, 2:26 PM
Genome Groups		ARWattam	3/26/15, 12:01 PM
M_haemophilum_new.fna	4.8 MB	ARWattam	9/30/15, 3:32 PM
M_haemophilum_new.txt	4.8 MB	ARWattam	9/30/15, 3:27 PM
Mycobacterium_marinum_PPE.faa	75.9 kB	ARWattam	11/5/15, 11:17 AM
NF2653.txt	3.2 MB	ARWattam	5/29/15, 2:15 PM
OanthropiATCC.txt	5.3 MB	ARWattam	8/3/15, 1:26 PM
Proteome Comparison		ARWattam	8/22/15, 10:18 AM
R2.txt	58.3 MB	ARWattam	7/13/15, 9:27 AM
SRR005751.fastq.gz	449.8 MB	ARWattam	11/13/15, 4:38 PM
SRR1019284.auto.contigs.fa	4.1 MB	ARWattam	7/13/15, 11:01 AM
SRR1033693.auto.contigs.fa	4.0 MB	ARWattam	7/13/15, 11:03 AM
bau_sim_R1.fq	58.3 MB	ARWattam	6/15/15, 11:11 AM

21. This will open a page that shows all the comparisons that you have in that folder. Click on the icon in front of the name of your job (Red arrow 1). This will open up the landing page for that job that shows you a diagram showing the sequence identity, a list of the names of the genomes in your job, and will also show a circular image that shows the relatedness. This will take a few seconds to load.



ARWattam / home / Proteome Comparison

	Size	Owner	Created
B_melitensis_comparison_16M_reference	6.3 kB	ARWattam	11/30/15, 9:32 AM
LES400 as reference	5.6 kB	ARWattam	10/28/15, 12:47 PM
M_marinum_as_reference	5.8 kB	ARWattam	10/22/15, 7:37 PM
leprae Br vs Refseq and Tn	5.0 kB	ARWattam	10/13/15, 7:18 PM

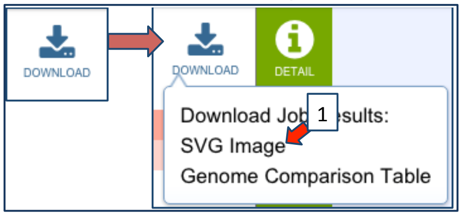
ARWattam / home / Proteome Comparison / B_melitensis_comparison_16M_reference

	Percent protein sequence identity															
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

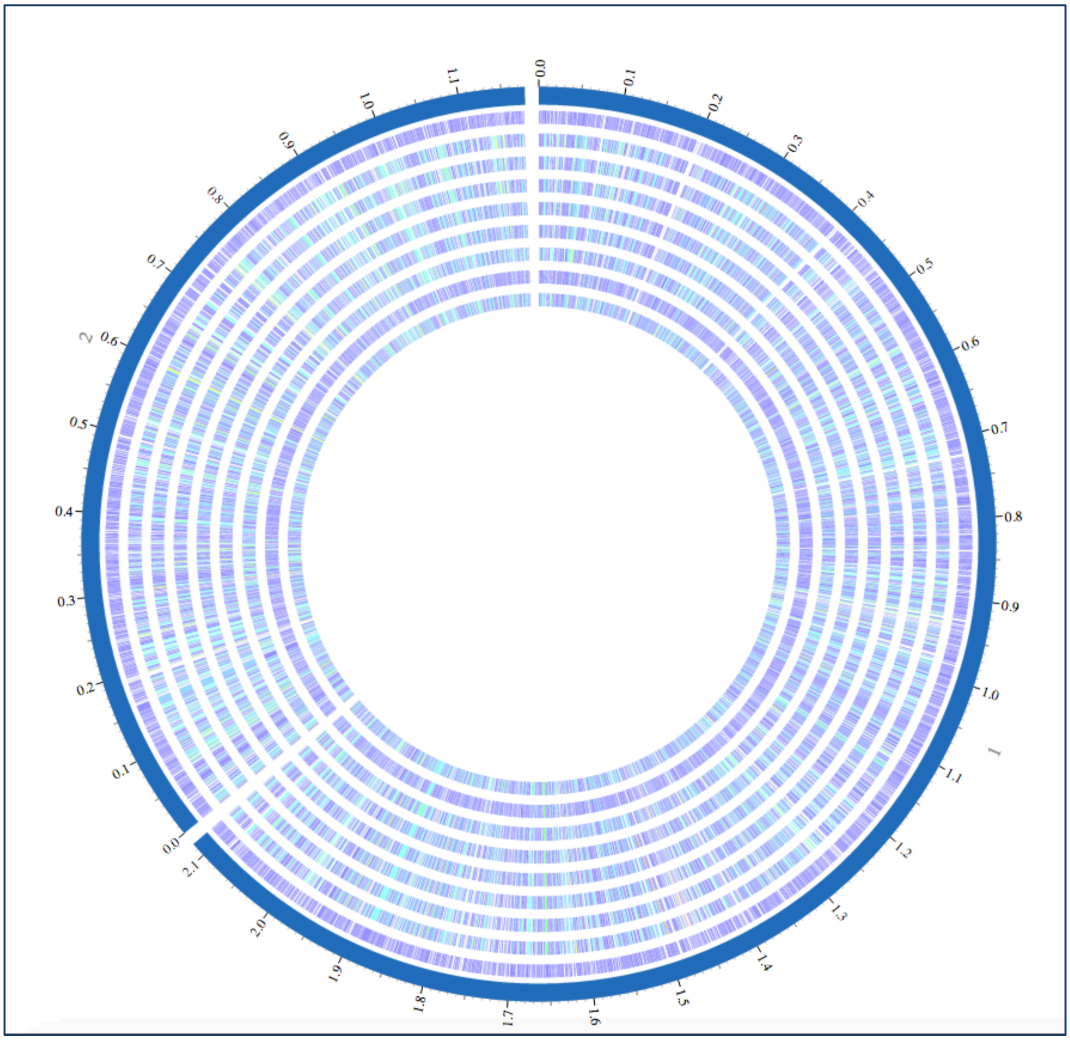
List of tracks, from outside to inside:

1. Brucella melitensis bv. 1 str. 16M (224914.11)
2. Brucella melitensis M5-90 (703352.4)
3. Brucella melitensis NI (1029825.4)
4. Brucella melitensis M28 (941967.3)
5. Brucella melitensis ATCC 23457 (546272.3)
6. Brucella melitensis bv. 2 str. 63/9 (520465.3)
7. Brucella melitensis bv. 3 str. Ether (520466.3)
8. Brucella melitensis bv. 1 str. Rev.1 (520464.3)
9. Brucella melitensis_F3_99_548_contigs (29459.27)

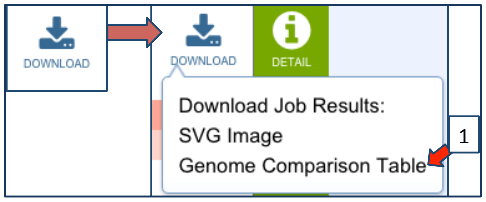
22. To see the entire image, find the download icon at the top of the page on the left and click on SVG image (Red arrow 1).



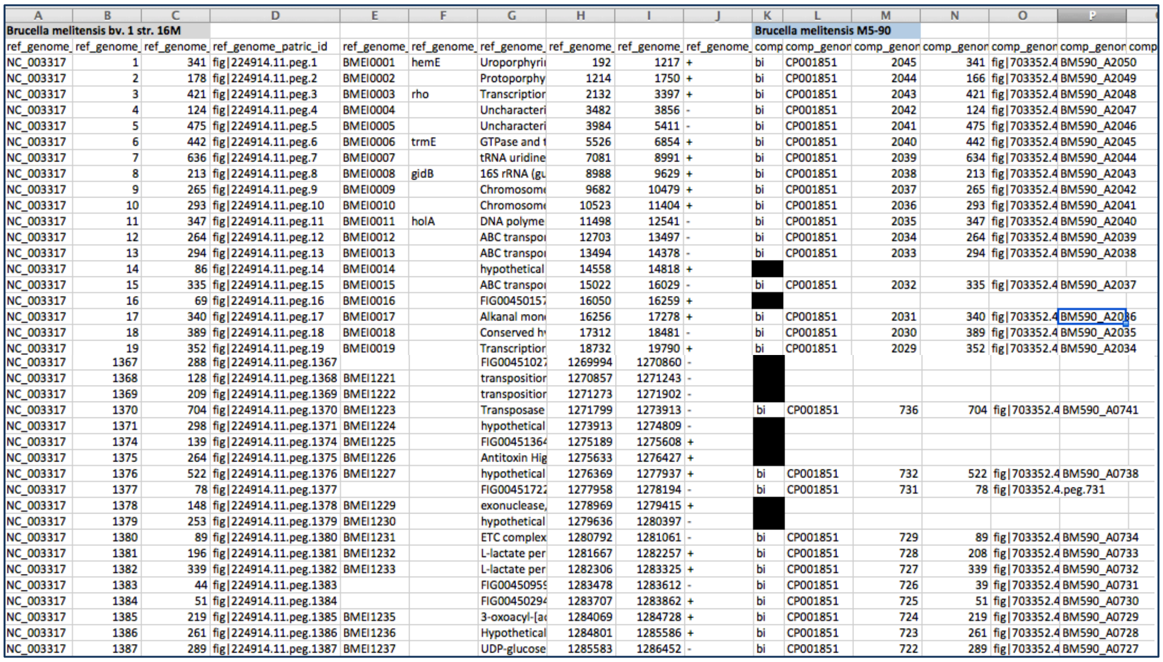
23. This will download the publication quality SVG image that shows the percent identity across all the proteins in the comparison genomes compared to the reference genome.



24. To examine the data underneath the visualization, click on Genome Comparison Table (Red arrow 1).



25. This will open up a text file, that you can put into excel. You will be able to see all the genes in the comparison genomes that have the best BLAST hits to the reference genome. There is a lot of information in this document. You should carefully examine all the column headings in row 2, as it will show you not only the name and chromosomal location of these genes, but also the size, the percent identity and sequence coverage.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
Brucella melitensis bv. 1 str. 16M										**Brucella melitensis M5-90**						
ref_genome_	ref_genome_	ref_genome_	ref_genome_patric_id	ref_genome_	ref_genome_	ref_genome_	ref_genome_	ref_genome_	ref_genome_	comp	comp_genor	comp_genor	comp_genor	comp_genor	comp_genor	comp
NC_003317	1	341	fig|224914.11.peg.1	BMEI0001	hemE	Uroporphyri	192	1217	+	bi	CP001851	2045	341	fig|703352.4	BM590_A2050	
NC_003317	2	178	fig|224914.11.peg.2	BMEI0002		Protoporphy	1214	1750	+	bi	CP001851	2044	166	fig|703352.4	BM590_A2049	
NC_003317	3	421	fig|224914.11.peg.3	BMEI0003	rho	Transcriptior	2132	3397	+	bi	CP001851	2043	421	fig|703352.4	BM590_A2048	
NC_003317	4	124	fig|224914.11.peg.4	BMEI0004		Uncharacteri	3482	3856	-	bi	CP001851	2042	124	fig|703352.4	BM590_A2047	
NC_003317	5	475	fig|224914.11.peg.5	BMEI0005		Uncharacteri	3984	5411	-	bi	CP001851	2041	475	fig|703352.4	BM590_A2046	
NC_003317	6	442	fig|224914.11.peg.6	BMEI0006	trmE	GTPase and t	5526	6854	+	bi	CP001851	2040	442	fig|703352.4	BM590_A2045	
NC_003317	7	636	fig|224914.11.peg.7	BMEI0007		tRNA uridine	7081	8991	+	bi	CP001851	2039	634	fig|703352.4	BM590_A2044	
NC_003317	8	213	fig|224914.11.peg.8	BMEI0008	gidB	16S rRNA (gu	8988	9629	+	bi	CP001851	2038	213	fig|703352.4	BM590_A2043	
NC_003317	9	265	fig|224914.11.peg.9	BMEI0009		Chromosome	9682	10479	+	bi	CP001851	2037	265	fig|703352.4	BM590_A2042	
NC_003317	10	293	fig|224914.11.peg.10	BMEI0010		Chromosome	10523	11404	+	bi	CP001851	2036	293	fig|703352.4	BM590_A2041	
NC_003317	11	347	fig|224914.11.peg.11	BMEI0011	holA	DNA polyme	11498	12541	-	bi	CP001851	2035	347	fig|703352.4	BM590_A2040	
NC_003317	12	264	fig|224914.11.peg.12	BMEI0012		ABC transpo	12703	13497	-	bi	CP001851	2034	264	fig|703352.4	BM590_A2039	
NC_003317	13	294	fig|224914.11.peg.13	BMEI0013		ABC transpo	13494	14378	-	bi	CP001851	2033	294	fig|703352.4	BM590_A2038	
NC_003317	14	86	fig|224914.11.peg.14	BMEI0014		hypothetical	14558	14818	+	■						
NC_003317	15	335	fig|224914.11.peg.15	BMEI0015		ABC transpo	15022	16029	-	bi	CP001851	2032	335	fig|703352.4	BM590_A2037	
NC_003317	16	69	fig|224914.11.peg.16	BMEI0016		FIG0045015	16050	16259	+	■						
NC_003317	17	340	fig|224914.11.peg.17	BMEI0017		Alkanal mon	16256	17278	+	bi	CP001851	2031	340	fig|703352.4	BM590_A2036	
NC_003317	18	389	fig|224914.11.peg.18	BMEI0018		Conserved h	17312	18481	-	bi	CP001851	2030	389	fig|703352.4	BM590_A2035	
NC_003317	19	352	fig|224914.11.peg.19	BMEI0019		Transcriptior	18732	19790	+	bi	CP001851	2029	352	fig|703352.4	BM590_A2034	
NC_003317	1367	288	fig|224914.11.peg.1367			FIG0045102	1269994	1270860	-	■						
NC_003317	1368	128	fig|224914.11.peg.1368	BMEI1221		transpositior	1270857	1271243	-	■						
NC_003317	1369	209	fig|224914.11.peg.1369	BMEI1222		transpositior	1271273	1271902	-	■						
NC_003317	1370	704	fig|224914.11.peg.1370	BMEI1223		Transposase	1271799	1273913	-	bi	CP001851	736	704	fig|703352.4	BM590_A0741	
NC_003317	1371	298	fig|224914.11.peg.1371	BMEI1224		hypothetical	1273913	1274809	-	■						
NC_003317	1374	139	fig|224914.11.peg.1374	BMEI1225		FIG0045136	1275189	1275608	+	■						
NC_003317	1375	264	fig|224914.11.peg.1375	BMEI1226		Antitoxin Hig	1275633	1276427	+	■						
NC_003317	1376	522	fig|224914.11.peg.1376	BMEI1227		hypothetical	1276369	1277937	+	bi	CP001851	732	522	fig|703352.4	BM590_A0738	
NC_003317	1377	78	fig|224914.11.peg.1377			FIG0045172	1277958	1278194	-	bi	CP001851	731	78	fig|703352.4.peg.731		
NC_003317	1378	148	fig|224914.11.peg.1378	BMEI1229		exonuclease,	1278969	1279415	+	■						
NC_003317	1379	253	fig|224914.11.peg.1379	BMEI1230		hypothetical	1279636	1280397	-	■						
NC_003317	1380	89	fig|224914.11.peg.1380	BMEI1231		ETC complex	1280792	1281061	-	bi	CP001851	729	89	fig|703352.4	BM590_A0734	
NC_003317	1381	196	fig|224914.11.peg.1381	BMEI1232		L-lactate per	1281667	1282257	+	bi	CP001851	728	208	fig|703352.4	BM590_A0733	
NC_003317	1382	339	fig|224914.11.peg.1382	BMEI1233		L-lactate per	1282306	1283325	+	bi	CP001851	727	339	fig|703352.4	BM590_A0732	
NC_003317	1383	44	fig|224914.11.peg.1383			FIG0045095	1283478	1283612	-	bi	CP001851	726	39	fig|703352.4	BM590_A0731	
NC_003317	1384	51	fig|224914.11.peg.1384			FIG0045029	1283707	1283862	+	bi	CP001851	725	51	fig|703352.4	BM590_A0730	
NC_003317	1385	219	fig|224914.11.peg.1385	BMEI1235		3-oxoacyl-[a	1284069	1284728	+	bi	CP001851	724	219	fig|703352.4	BM590_A0729	
NC_003317	1386	261	fig|224914.11.peg.1386	BMEI1236		Hypothetical	1284801	1285586	+	bi	CP001851	723	261	fig|703352.4	BM590_A0728	
NC_003317	1387	289	fig|224914.11.peg.1387	BMEI1237		UDP-glucose	1285583	1286452	-	bi	CP001851	722	289	fig|703352.4	BM590_A0727	