

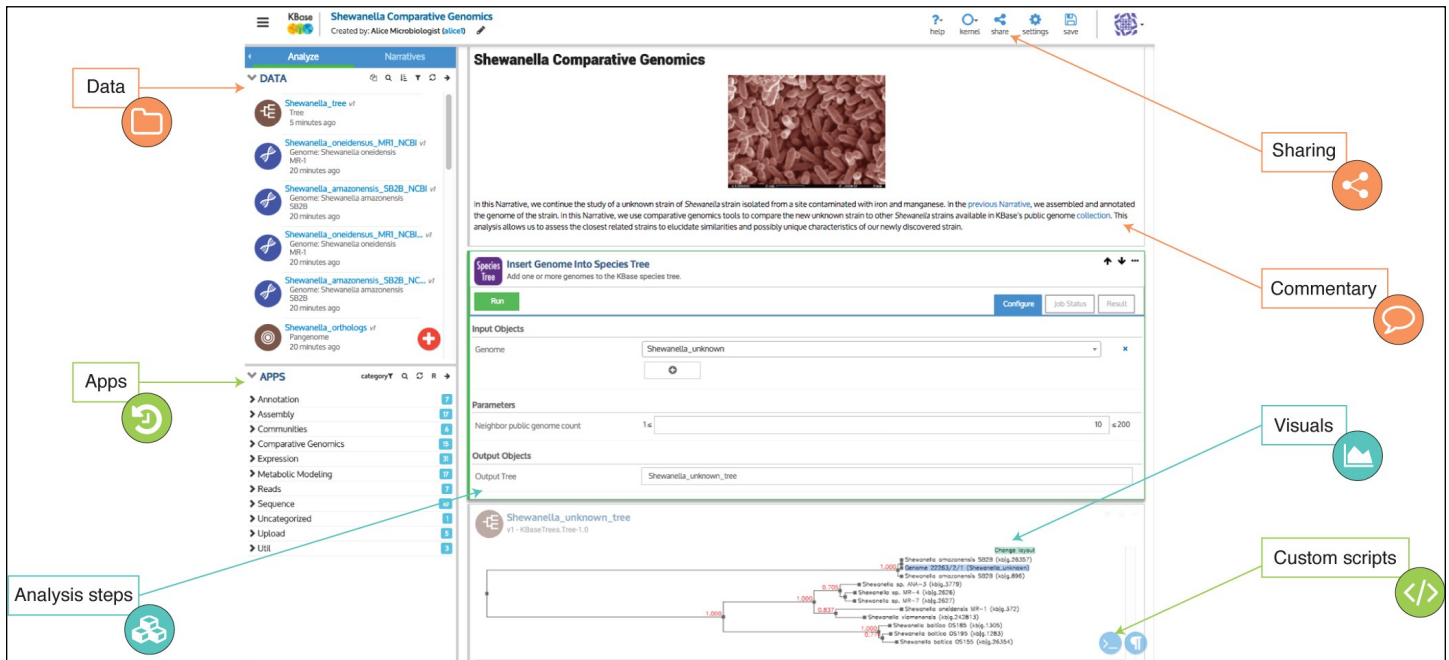
The United States Department of Energy Systems Biology Knowledgebase

What is KBase?

Open-source software and data platform with an graphical interface that enables data sharing and analysis of microbes and microbial communities...

To use KBase you can login/register through a google account.

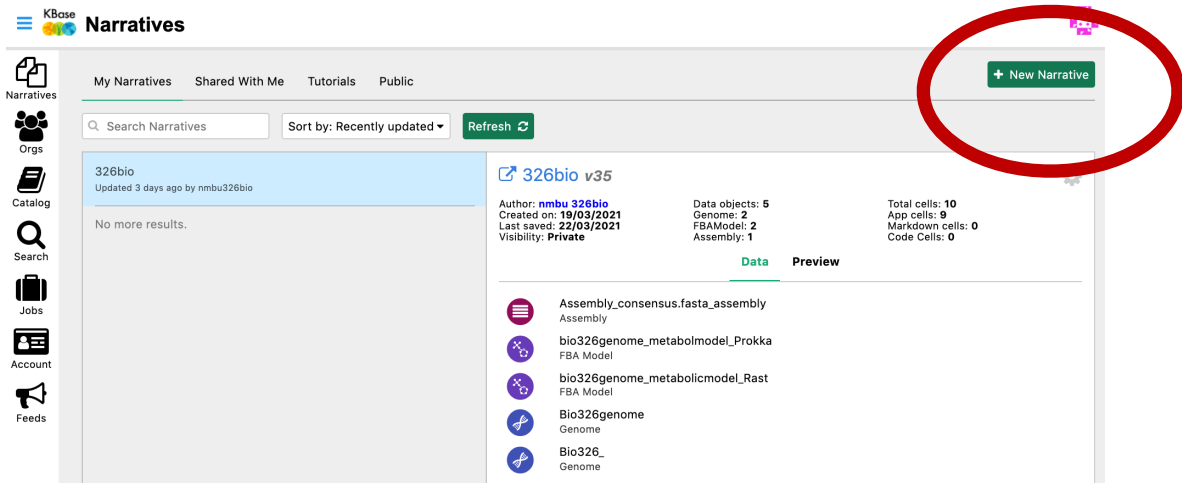
Overall look of KBase:



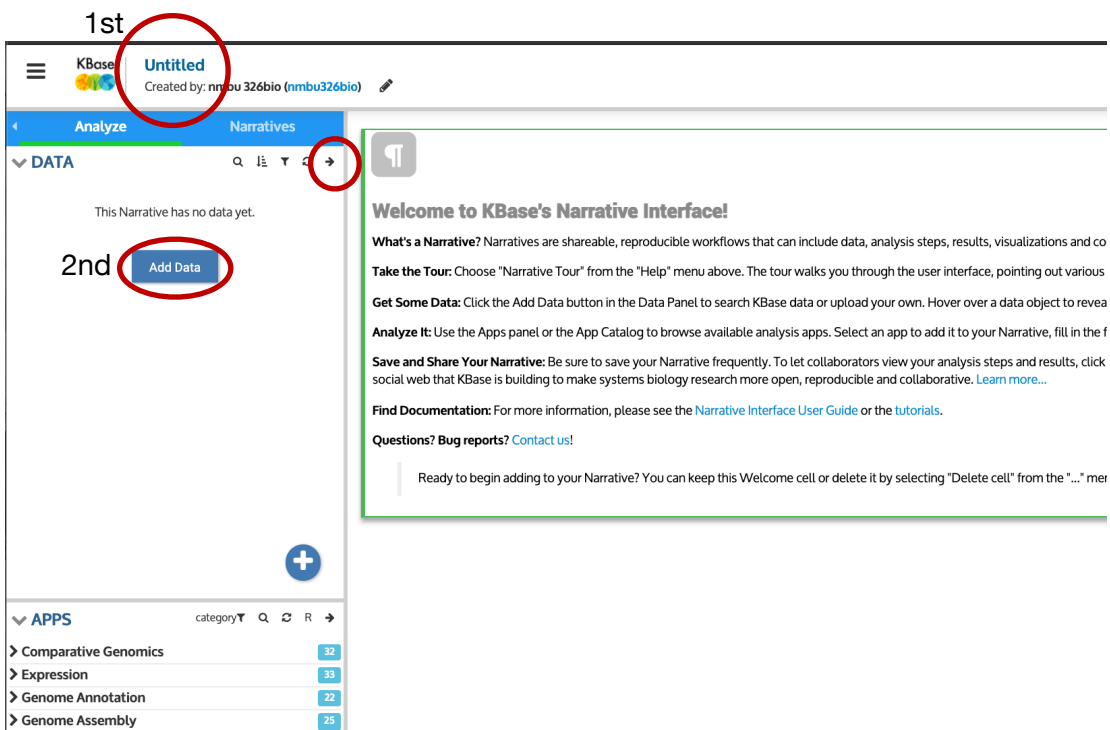
The screenshot displays the KBase interface for a 'Shewanella Comparative Genomics' workflow. The left sidebar shows a list of data items and applications. The main workspace contains a species tree visualization and a workflow diagram. The right sidebar offers options for sharing, commentary, visuals, and custom scripts. Arrows from labels point to these specific UI elements.

When you first start, you will need to make a “Narrative”.

Narrative: is a “notebook” that holds your computational experiment.



It will load a new Narrative that is automatically called Untitled. First thing you should do is rename it, you can do this by clicking “Untitled”



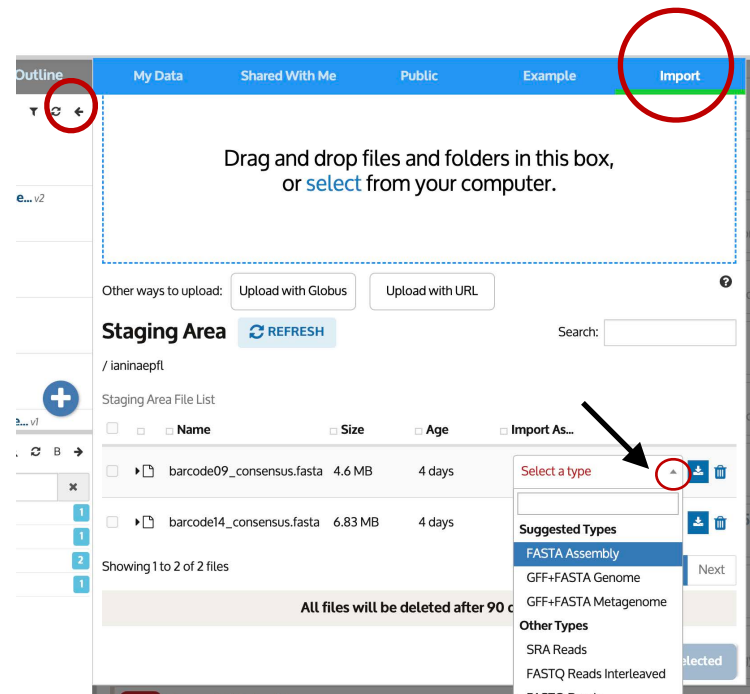
Getting you data on KBase

You are directed to the import tab. In general, you need to first upload the files to your account, and then import into your specific Narrative.

I have the *fasta* file

“Assembly_consensus.fasta” of your assembly ready to go! Don't worry the .fa and .fasta files are interchangeable

Once uploaded, the file will go into the “Staging Area”.



To import the assembly into your Narrative, you will need to scroll down to get the “FASTA Assembly” option and then hit the ‘blue up arrow button’.

Once data is uploaded you can use the Apps

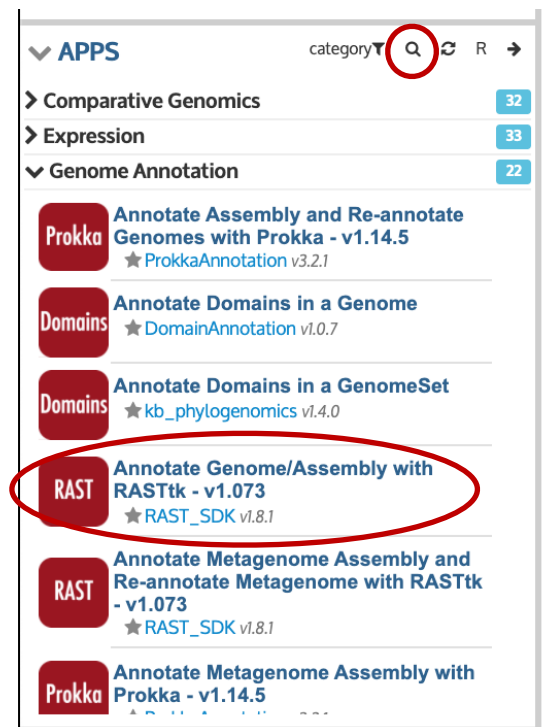
Lots of Apps including ones for:

Genome annotation: **RAST**, **Prokka**, **DRAM** you can check how the results differ when annotating with these different methods.

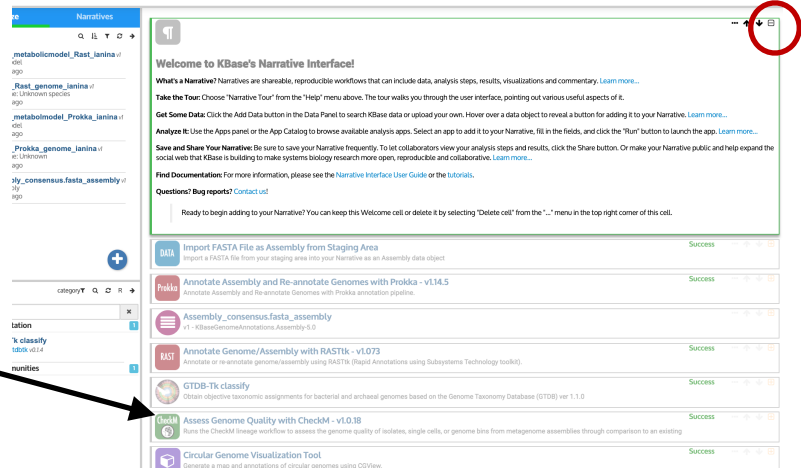
Genome assessment: **GTDB-tk**, **CheckM**

Metabolic Modeling: you can **Build Metabolic Model** with with the assembly/genome you annotated with either RAST or Prokka. Why not check how the different annotations effect the metabolic modelling results?

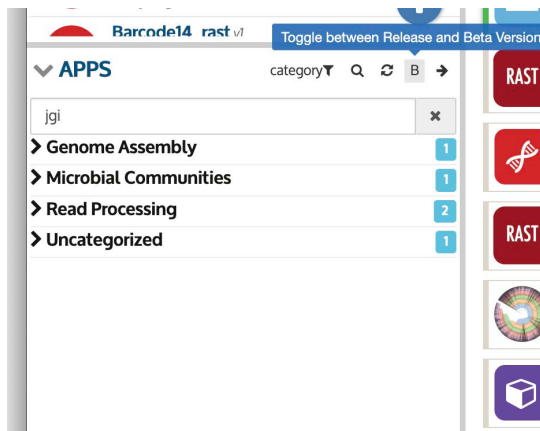
Also Apps for: (meta)genome assembly, Contig binning, Community modelling, etc.....



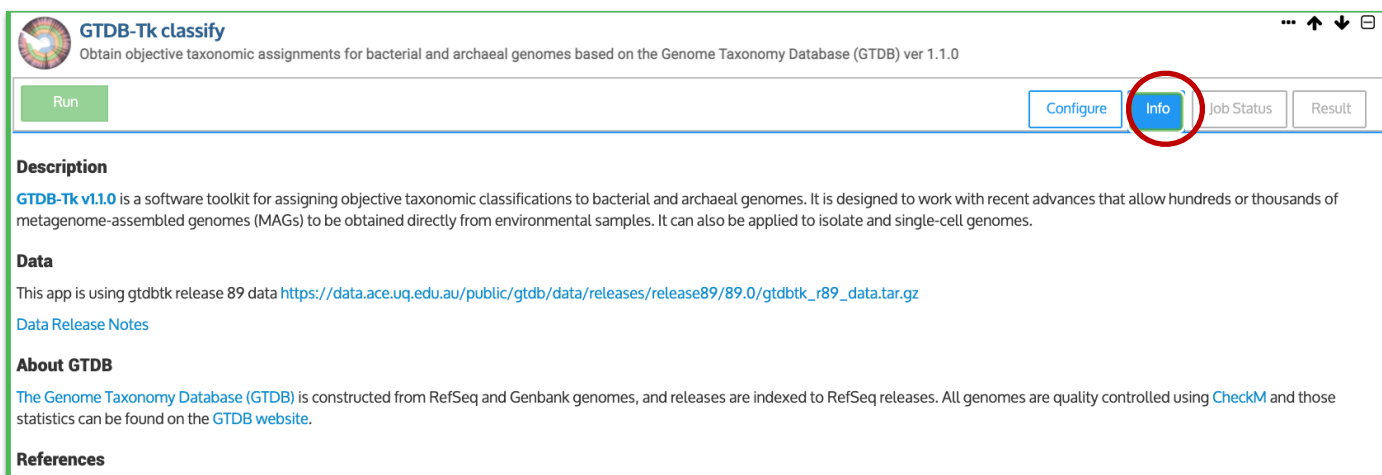
Every action/App performed or initiated by a user appears as a 'cell' in the Narrative. You can expand the cells by clicking on the + and - signs on the very right of the cell.



You can look for Apps in the bottom left corner of KBase. Once you click on an App it will appear as a cell in the Narrative. Once in the cell, each App will have several tabs on it. You can click the “Info” tab to see what the App does and have access to additional references. The “Configure” tab is where you can select what data/file you want to use the App with. In your case it would likely be your imported assembly file or the annotated genome. Sometimes you can also select other options in the “Configure”, I recommend reading about the App and just trying a few things. Not all Apps will work with all data types. Click “Run” button to start the analysis.



Switch to be able to see the apps that are still in Beta mode.



Results

Once the App is finished running, the results are either in the cell under the “Results” tab of the App you ran, or they are created as objects in the ‘Data’ panels. If you object is in the Data panel you can click on it to explore the data in the cell, or on the ‘binocular’ looking button to explore the results in a separate window. Try checking what the other buttons do.

This is example of output from the GTDB app that can taxonomically classify your organism.

Note: To run GTDB you need to first convert your assembly into an AssemblySet.

The screenshot shows the GTDB-Tk classify app interface. The 'Report' tab is active, displaying a table of taxonomic classifications. A red circle highlights the 'View report in separate window' button. The table has columns for User Genome, Classification, FastANI Reference, FastANI Radius, FastANI Taxonomy, FastANI ANI, FastANI Alignment Fraction, and Closest Placement Reference. The data row shows the classification of the assembly 'Assembly_consensus.fasta Assembly' as d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Psychrobacter; s__Psychrobacter sp000247495.

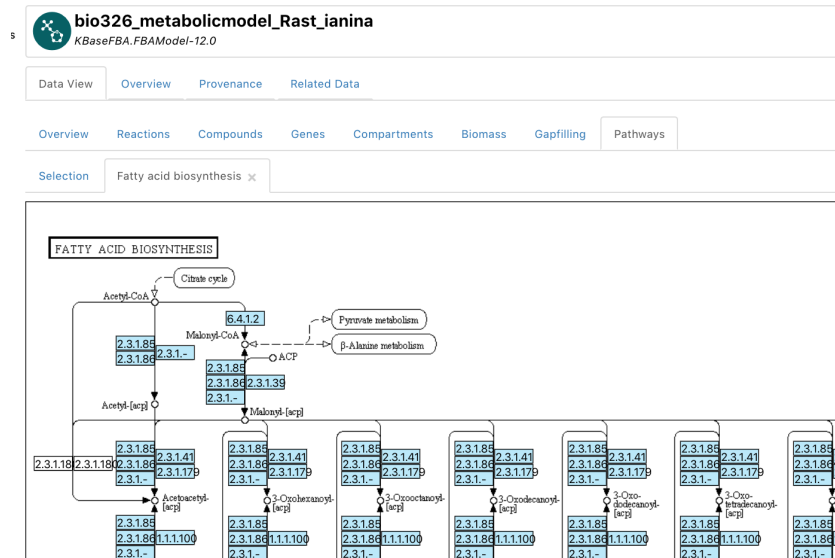
User Genome	Classification	FastANI Reference	FastANI Radius	FastANI Taxonomy	FastANI ANI	FastANI Alignment Fraction	Closest Placement Reference
Assembly_consensus.fasta Assembly	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Psychrobacter; s__Psychrobacter sp000247495	GCF_000247495.1	95	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Psychrobacter; s__Psychrobacter sp000247495	95.97	0.84	GCF_000247495

Here is example output from annotation of the assembly/genome with RAST. You need to scroll down to get to the Assembly and Annotation part.

The screenshot shows the RAST genome annotation interface. The 'Assembly and Annotation' tab is active, displaying a list of functional categories on the left and a table of gene annotations on the right. A red circle highlights the 'Assembly and Annotation' tab. The table has columns for Gene ID, Contig, Start, Strand, Length, Type, and Function. The data row shows the annotation of the gene 'Bio326_CDS.2' as a Putative von Willebrand factor, vWF type A domain protein STM2315.

Gene ID	Contig	Start	Strand	Length	Type	Function
Bio326_CDS.2	tig000000001	217	+	525	gene	Putative von Willebrand factor, vWF type A domain protein STM2315
Bio326_CDS.3	tig000000001	816	+	216	gene	-
Bio326_CDS.4	tig000000001	1392	+	273	gene	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)
Bio326_CDS.5	tig000000001	1661	+	1461	gene	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)
Bio326_CDS.5621	tig000000001	3202	+	249	gene	-
Bio326_CDS.6	tig000000001	3618	+	1227	gene	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)
Bio326_CDS.7	tig000000001	4805	+	300	gene	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)
Bio326_CDS.8	tig000000001	5101	+	156	gene	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)
Bio326_CDS.9	tig000000001	5249	+	150	gene	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)
Bio326_CDS.10	tig000000001	5658	+	450	gene	Ribonuclease PH (EC 2.7.7.56)

Here is example output from the Build Metabolic Model app using the assembly/genome that was annotated with RAST as an input for this app.



If you get stuck/lost KBase has a useful help section.

Have fun **exploring the Apps** with your data for the report!

