The purpose of this document is to provide information through the frequently asked questions (FAQ) from our GalaxyTrakr users. Please review this FAQ document to learn the basics of GalaxyTrakr.

1. **What is GalaxyTrakr, and how is it different from GenomeTrakr?**
   a. GalaxyTrakr is GenomeTrakr's own hosted instance of the Galaxy Project, an open source, web-based platform for collaboration and science. GalaxyTrakr's team of engineers and developers curate tools for use in workflows and data exploration by the community of food pathogen researchers.
   b. GenomeTrakr is an international network for WGS biosurveillance of foodborne pathogens, the first of its kind and the most open. GenomeTrakr supports the investigation and resolution of outbreaks of foodborne illness by providing an up-to-date view of pathogen population structure combined with source collection metadata.

2. **How do I create a GalaxyTrakr account?**
   a. If you would like to create a GalaxyTrakr account, please visit: https://account.galaxytrakr.org/Account/Register, to register. Once registered, you will have to verify your email address so that we can approve your account for GalaxyTrakr use.

3. **I forgot my username and/or password, is it possible to retrieve both?**
   a. Yes, please contact GalaxyTrakr Support (GalaxyTrakrSupport@fda.hhs.gov) for account related issues.

4. **I am running into a blank screen after I log in. What am I doing wrong?**
   a. GalaxyTrakr supports browsers including Google Chrome, Mozilla Firefox, Apple Safari, Microsoft Edge, and others. Internet Explorer 11 and earlier may not be supported. If problems persist, try clearing your browser's cache and then login again.

5. **When I log in again, I get a message that says, "No session token set, denying request." How do I log in?**
   a. This is a proxy server error on our end. Until we can fix it, edit the URL by deleting `user/login` and everything following, then hit enter. You should be redirected to the login page.

6. There are a handful of different CFSAN SNP Pipeline workflows available on GalaxyTrakr, what are the differences between the different versions?
    a. Lite – only outputs the filtered SNP files
    b. w/ Assembly – does skesa assemblies for all samples and picks the best assembly as the reference
    c. w/ SRA download and assembly – downloads reads from NCBI, assembles with Spades, and picks the best assembly as reference
    d. Since Galaxy lets you use entire workflows as steps in other workflows, the Lite versions of the workflow are better-suited to being steps in any other workflows you might create since they have fewer outputs. For the most part, the full workflows are presented only for the sake of completeness.

7. In the "Shared Data", there are a few CFSAN SNP Pipelines that have the same name or a slight variation in the name. Does it matter which one we import and use? Is there a favorite amongst the GalaxyTrakr team?
    a. Justin leans pretty heavily on the "CFSAN SNP Pipeline w/ Assembly Lite" – try it as a sub-step in your own workflows.
    b. The "core" SNP Pipeline comes in normal and "lite" flavors, where the normal pipeline returns two sets of results – one each for filtered and unfiltered SNP sets. The Lite pipeline returns results for only the filtered set. We extend the core with an assembly and reference choosing step – and extend that with an SRA download step – that's "w/ SRA Download and Assembly". In practice, the difference is what you're required to provide as input.
        i. The core pipelines require paired-end reads and a reference assembly
        ii. The "w/ Assembly" pipelines require only the paired-end reads (because they're all assembled and then the "best" assembly is chosen for you by the pipeline). It does skesa assemblies for all samples and picks the best assembly as reference
        iii. The "w/ SRA Download" pipelines only require a text file with SRA accession numbers. It downloads reads from NCBI, assembles with Spades, and picks the best assembly as reference
    c. Galaxy lets you use entire workflows as steps in other workflows, the Lite versions of the workflow are probably better-suited to being steps in any other workflows you might create since they have fewer outputs. For the most part, the Full workflows are presented only for sake of completeness; the Lite workflows are the ones most people wind up using.

8. Is there a limit for number of isolates a user can run in one time?
    a. There aren't any hard limits except those imposed by the size of your disk space quota; in practice that probably limits the pipeline to about 200 samples or so. It's possible for us to expand yours when there's a useful research purpose. Other than that, the limit could be in the thousands. The SNP Pipeline, for example, has been run on some pretty large datasets.

9. How do I delete my history?

    a. To purge individual datasets from a history, switch to the history that contains the data and click "Operations on multiple datasets", check all datasets you want to delete and then click "Permanently delete datasets". This will delete and purge (permanently remove) the datasets selected.

    b. If all of the datasets in a history are no longer needed then the entire history should be purged. Click on "View All Histories". Click on the drop-down arrow for the history, then click on Purge. This will permanently delete the history.

    c. When deleting an entire history that contains a large amount of datasets it may take more than 30 seconds to complete. If it times out, please wait at least an hour and then come back to the "View All Histories" page to see if the history still exists. The purge process happens in the background and does not always finish prior to the time out period of the web interface.

10. There is the "NGS: CFSAN SNP Pipeline (beta)" in the NGS toolbox. How different is that one compared to importing one of the CFSAN SNP Pipelines?

    a. In order to build the workflows, we had to wrap the components of the SNP pipeline in Galaxy tool wrappers, so they appear as tools in the tool panel. They're the same tools being used by the workflow, but you have to orchestrate them yourself if you use them from the tool panel, and there's a little bit of dataset renaming that goes on in the workflow that the tools don't do on their own. So, it's easy to run off the rails if you just use the pipeline stages on their own from the tool panel, which is why we recommend the workflows instead. We considered wrapping the whole pipeline into a single Galaxy tool, but then it won't parallelize correctly over the compute cluster so it would take forever to run. As separate tools, we can lean on Galaxy's job orchestration tools and balance the various loads across the compute resources. But it does lead to a little bit of user confusion, unfortunately.

11. Are workflows sensitive to input file names and special characters?

    a. Yes, workflows can be sensitive to input files that have dashes, spaces, and underscores in the dataset name, and will throw an error if the dataset has certain characters in the name. Any character you wouldn't use as a filename in a command-line shell should probably be excluded from your dataset and sample names, if possible.

    b. MicroRunQC: Underscores, dashes and spaces are all acceptable for naming dataset pairs. From our experience, having these characters allows for a successful execution.

    c. SeqSero2v1.0.2: Underscores and acceptable.

    d. CFSAN SNP Pipeline w/ Assembly Lite: Spaces are not allowed – QUAST changes them in its output files and this breaks the pipeline at the reference selection stage. (See the next FAQ.)

12. What should I name my dataset pairs/samples when running the CFSAN SNP Pipeline w/ Assembly?

    a. For the most part, the restrictions on what a user can use for a sample name in the SNP Pipeline workflow are related to QUAST's behavior with sample names, which is to replace several different ASCII characters or omit them entirely. This fouls up the next step of the workflow, which is to look up the sample name of the best-performing assembly according to QUAST and retrieve it from the set of assemblies by its sample name. Since the name is changed, the assembly in question can't be found and the pipeline fails.

    So, with that in mind, when going through and trying various combinations of ASCII characters in names to see what QUAST will change, so far:

        i. Spaces, ( ), period and comma, *, ^, @, #, : -> underscores
        ii. [, {,], } -> "ob", "oc", "cb", "cc" ("open/closed" "bracket/curly brace")
        iii. |, `,~,&, %, $,: -> "X"
        iv. Single quote -> "sq"

    Dashes and underscores, surprisingly, are fine – they're not changed by QUAST at all.

    In general:
    CFSAN000123_01-4 -> should work as a sample name
    CFSAN000123.456: -> don't even try

13. It's fun to explore the different tools but now I'm working on a huge dataset. How do I collate and download my results?

    a. Galaxy bundles up multiple downloads as a tar.gz file; you'll need to decompress it before you can access the files. For tabular datasets, you can merge multiple results while keeping a single table header using the Merge Tabular Datasets tool under Join, Subtract, and Group.