# Source variability protocol (written by: Timothy Lim, version 30/10/2024)

This document provides a guide on how to run the entire process to determine if there are enough samples within the faecal source library to accurately represent the faecal sources, using SourceTracker.

1. Prepare metadata file that only contain sources.
2. Open “*Create\_metadata\_source\_variability.py*”, then do the following:
   1. Change *metadata\_path* and *save\_path* to the stated metadata file in step (1) and your desired folder of saving modified metadata
   2. Run the script
   3. You should see that a lot of *txt* files have been saved in under *save\_path*
3. To write SourceTracker script, open “*ST\_sinkswitch\_16S.py*” and do the following:
   1. Change *files* and *namefile* to the folder where you saved your modified metadata as per step (2)
   2. Under *script*
      1. Change *-i* to your feature table
      2. Change *-m* to your modified metadata folder
      3. Change *-o* to your SourceTracker output folder
   3. Run the script
   4. You should see that *ST\_sinkswitch\_16S.txt* has been saved to the same directory as “*ST\_sinkswitch\_16S.py*”
   5. Upload the *ST\_sinkswitch\_16S.txt* to<https://www.fileformat.info/convert/text/dos2unix.tr> to solve \n issue, and download resulting *txt* file (split into two or more files if error says that the file is too long)
   6. Change the file type of the resulting *txt* file to *.sh*
   7. Add #!/bin/bash to the top row of the resulting .sh file (open using NotePad++ or something else)
4. Login to WinSCP and PUTTY, run SourceTracker to run the .sh files. Check if there are any two lines of script that were “merged” together (as a result of glitch), and put them into two different rows where necessary.
5. To create metadata for compilation (or check for missing runs), open “*ST\_sinkswitch\_16S\_source\_variability.py*”
   1. Change *files* and *namefile* to the folder where you saved your modified metadata as per step (2)
   2. Run the script
   3. You will receive a *ST\_sinkswitch\_16S\_Source.txt* file with metadata entries such as *Waterbird\_5Sources\_DrawNo1.*
   4. Modify this metadata (i.e., duplicate the file, and rename, let’s say *ST\_sinkswitch\_16S\_Source\_modified.txt*), such that you have 3 columns, *#SampleID, Env,* and *SourceSink*. For *#SampleID* and *Env*, you can just copy paste the metadata entries, with *SourceSink* being just *source*.
6. To check for any missing files (due to Massive session expiry), open *“Identify SourceTracker outputs with missing files.py”*
   1. Change the *directory* to the SourceTracker export (actual path to the directory you want to search)
   2. Run the script
   3. This script will create a *missing\_files.txt* file in the current working directory, containing the paths of the folders that do not have the *mixing\_proportions.txt* file
   4. Delete the folders which do not have the *mixing\_proportions.txt* file from WinSCP
   5. Re-run SourceTracker
7. To compile the results, open *“Compilation\_SourceVariability\_withFolderName.py”*
   1. Change *files\_path* to your SourceTracker\Exports folder
   2. Change *env\_file\_path* to *ST\_sinkswitch\_16S\_Source\_modified.txt* (or any other file as per step 5d
   3. Change save\_path to a new *Results* folder for Source variability
8. Sort the resulting file according to the sample ID (ascending), and calculate the relative standard deviation (RSD).
9. Compile all the replicate runs using “OFFSET” function (see *Filter1* for example)
10. Duplicate *Filter1* (named as *Filter2*), then identify number of sources and the animal.
11. Duplicate *Filter2* to another file, and then filter according to each animal. Make a PivotTable, and use the following fields to get the summary result for each animal.

A screenshot of a computer

Description automatically generated

1. Calculate the number of correctly predicted samples against total number of cases.