

User manual for generating TRAMS input files from vcf and Mauve output files

Licence terms: These scripts are provided free to academic users, allowing them to use, share or modify them but any commercial use/exploitation will require a written consent from the authors.

Disclaimer: The program/scripts have been tested on a variety of datasets and works well. However, the authors provide no guarantees/warranties about the performance, fitness for a particular purpose, or any other warranties whether expressed or implied. The authors take no responsibility and are not liable for direct, indirect, special, incidental, or consequential damages resulting from the use, misuse, or inability to use this software under any circumstances.

Contact: Vartul Sangal (vartul.sangal@strath.ac.uk); Paul A. Hoskisson (paul.hoskisson@strath.ac.uk)

Bug reports: Richard A. Reumerman (richard.reumerman@strath.ac.uk)

Citation: Reumerman RA, Tucker NP, Herron PR, Hoskisson PA, Sangal V. Tool for Rapid Annotation of Microbial SNPs (TRAMS): a simple program for rapid annotation of genomic variation in prokaryotes, *Antonie van Leeuwenhoek Journal of Microbiology* (*submitted*)

1. Conversion of Mauve SNP table into TRAMS input: *mauve2trams.py*

Mauve SNP output file need reformatting for TRAMS annotation. We provide a script, *mauve2trams.py*, to convert Mauve format to the TRAMS format. The strain order in the Mauve output depends on their original input order for the alignment but users may want to select a different reference for annotation that is not the first in the alignment. This script provides users this opportunity to select any of the strains to be the reference and accordingly generates the input file excluding the loci absent in the selected reference.

The users can simply run the script by typing **mauve2trams.py** on the command line that will prompt to enter:

1. Name of the input file
2. Name of the output file

The script will now sort all the sequences in the input file and will prompt to select one of them as reference

3. Number of the reference strain from the list

Alternatively, users can provide all the arguments in a single command line:

```
mauve2trams.py -i input.txt -o output.txt -r Reference
```

```
-i      name of the input file
-o      name of the output file
-r      name of the reference sequences from the Mauve input file
```

This script is also provided as **executable files** for 32-bit and 64-bit WINDOWS users (***mauve2trams.exe***) and will prompt to enter the values on the screen as described above.

2. Conversion of vcf file into TRAMS input: *vcf2trams.py*

Some mapping programs generate SNP data in variant call format (vcf) that need reformatting for annotation by TRAMS. We provide a python script, *vcf2trams.py*, for converting SNP tables in vcf format into TRAMS input files. *vcf2trams.py* will need a SNP file in vcf format and a user defined

quality cut-off for selecting the loci. Please note that this script is designed to handle vcf files containing SNPs between a reference and only one variant strain.

The users can simply run the script by typing **vcf2trams.py** on the command line that will prompt to enter:

1. Name of the input file
2. Name of the output file
3. Name of the reference strain
4. Name of the variant strain and
5. A quality cut-off value.

Alternatively, users can provide all the arguments in a single command line:

vcf2trams.py -i input.vcf -o output.txt -r Reference -a Variant -t 10

-i name of the input file
-o name of the output file
-r name of the reference strain (default Ref)
-a name of the variant strain (default Alt)
-t cut-off quality score for SNP selection (default 0)

If name of reference and variant strains are not defined, the program will use Ref and Alt as default names for reference and variant strains, respectively. VCF files contain quality scores for each SNP (higher score means higher certainty) and user can define a cut-off value for SNP filtration. The program will include all SNPs from the file if no cut-off is defined.

This script is also provided as **executable files** for 32-bit and 64-bit WINDOWS users (**vcf2trams.exe**) and will prompt to enter these values on the screen. Again, if user does not provide names of reference and variant strains and do not define a cut-off quality score the default values will be used as mentioned before.

3. Merging multiple TRAMS input files: *tramstablemerger.py*

This script is designed to merge different TRAMS input files into one. This is especially useful to combine tables independently obtained by converting different formats (vcf and Mauve to TRAMS) using the same reference.

The users can simply run the script by typing **tramstablemerger.py** on the command line that will prompt to enter the name of the output file followed by the input file names. The script only accepts one file at once and users need to push return (enter) key after each file name. When all file names are entered, the users will need to press the return key twice and an output file will be generated.

If the reference base for a certain position varies in different files, the program will print an error message in the output file and on screen and will exit. **Always check the output file to see if the last line is not an error message.**

This script is also provided as **executable files** for 32-bit and 64-bit WINDOWS users (**tramstablemerger.exe**) and will prompt to enter the file names as described above.