

1 Compare-all.py: A Comprehensive Guide

1.1 Background

Compare-all is a python command line tool within the krepe package that allows bioinformaticists to count the kmers with any number of .fasta, .fna, or .fastq files, and also lets the user input any number of files which are then parsed into kmers of length k and then visualised using a dendrogram.

1.2 Installation:

Compare-all is a one of KREPE's many python command line tools for bioinformaticists. Organism-comparison makes use of the powerful sour-mash python library to make basic k-mer analysis easier than ever. There are two ways to install KREPE.

- pip install krepe
- git clone <https://RGBwalnut/KREPE.git>

1.3 Usage

If it was installed with pip you can use:

```
$ compare-all (k-mer length) (file-1) (file-2) (file-3) (file-n) (file type) (labels or -base)
```

Or if it was installed from github you can:

```
$ python3 PATH/compare-all.py (k-mer length) (file-1) (file-2) (file-3) (file-n) (file type) (labels or -base)
```

1.4 Tutorial

If you installed pip then navigate to your data directory and run

- `ls -d $PWD/* > soil_data.txt`
- `compare-all.py 10 'cat soil_data.txt' -fastq -base`

But if you git cloned then you can run in your data directory:

- `/KREPE/compare-all.py 10 'ls *.fastq' -fastq -base`

1.5 Results

You should get a dendrogram and color matrix that shows how similar each sample is when compared to every other sample. The dendrogram will then cluster the samples based on clade.

