

1 Organism-comparison: A Comprehensive Guide

1.1 Explanation:

Organism-comparison allows the user to count k-mers of a specified k-mer length and create a dictionary holding that k-mer and its abundance. It then allows the user to compare two DNA samples whether they both be .fna, .fasta, or .fastq. Once the k-mer dictionary of both files has been created, organism-comparison will calculate the Jaccard similarity of the files. This can be used to calculate the genetic diversity of the two organisms. Organism-comparison allows you to then create a venn diagram of the intersection of the k-mer dictionaries.

1.2 Installation:

Organism-comparison 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:

```
$ organism-comparison (k-mer length) (file1) (file2) (file type)
```

Or if it was installed from github you can:

```
$ python3 PATH/organism-comparison.py (k-mer length) (file1) (file 2) (file type)
```

1.4 Tutorial:

If you are using our pip package you can simply run the program with the amazon.fastq and permafrost.fastq file in the same directory.

- \$ organism-comparison 21 amazon.fastq permafrost.fastq -fastq

But if you are using the github version of our program you will need to cd into the comparison tutorial directory and give the absolute path for organism-comparison.py.

- `$ cd /KREPE/tex-tutorials/comparison-tutorial/`
- `$ python3 /KREPE/tex-tutorials/comparison-tutorial/organism-comparison.py
21 amazon.fastq permafrost.fastq -fastq`

1.5 Results:

This will yield a venn diagram that shows the set of each list of k-mers that occur in the files with the overlap being kmers that they share.

