CompàreGenome

User Guide to installation and use of **CompàreGenome**

WHAT IS

CompàreGenome is a command-line tool to perform genomes comparative analysis of different organisms, both eukaryotes and prokaryotes, by Whole Genome Sequences analysis.

CompàreGenome identifies with high level of precision and efficiency the fine genetic differences existing between genomes, detecting the most subtle genetic variations, resulting in ready-to-use information for users.

REQUIREMENTS

- Anaconda/Bioconda
- Unix-like systems
- Reference genome (in Genebank format)
- Query genomes (In fasta format)

STEPS

1. PRELIMINARY SETTINGS

1.1. CompàreGenome requires *Anaconda*. If not already available, please download and install the latest version. Then, update and initiate your Anaconda version by the following commands (based on version 22.9.0):

```
$ conda update -n base -c defaults conda    #to update. Please close and restart your shell after that
$ conda init      #to initiate. Please close and restart your shell after that
```

1.2. Either download **CompàreGenome** at https://bioecopest.com/comparegenome/ or clone the repository at GitHub:

```
$ git clone https://github.com/gmoro-bioecopest/CompareGenome.git
```

2. INSTALLATION

2.1. Once you have downloaded and unzipped the file, navigate into the CompareGenome directory:

```
$ cd CompareGenome
```

2.2. Run the following command to install (default value for the number of cores is set to half of the total number of cores in the system:

```
$ bash install.sh -n [Number of cores to use. ]
```

After the installation process completes, the tool will be available in the current directory and can be executed with the command: `./comp-gene.sh`

NOTE. *The installation may take several minutes.*

CompàreGenome

User Guide to installation and use of **CompàreGenome**

ARGUMENTS

```
Usage: comp-gene.sh          [--num-cores/-n <num_cores>] [--reference/-r <reference_file>]
                    [--genomes/-g <list of genomes to compare>]
                    [--output-dir/-o <output_folder>]


Run CompareGenome with the specified options.


Options:

--num-cores, -n  Number of core to use. Default: 4
--reference, -r  The full path of the reference genome, in genebank format.
                 You can use your own reference file or alternatively download from public databases.
                 It is best that the reference genome and your genome assemblies belong to the same species.
                 In case this is not possible, you can choose the most closely related species.
--genomes,   -g  The full path of the genome assemblies to compare.
                 The files must be in fasta format and at least 2 genomes are required.
                 If you have one genome assembly only, you can download the missing genome assembly from public databases.
                 In this case it is best to choose the same species of your own assembly, or a closely related one.
--output-dir,-o  The full path of the output folder. CompareGenome will save analysis outputs in this folder.
--help, -h       Show this help message and exit
```

- Show help:

```
$ ./comp-gene.sh -h
```

EXAMPLES

NOTE. *before starting, be sure that no conda environment is already activated*

INSTALLATION

From the unzipped CompàreGenome folder (in the example CompareGenomeV2.1test), run install.sh by indicating the number of cores in your machine (in the example 8).

```
% cd /yourpath/CompareGenomeV2.1test
-CompareGenomeV2.1test %  bash install.sh -n 8
```

ANALYSIS

Comparison of 2 genome assemblies (genome1.fa and genome2.fa)

```
-CompareGenomeV2.1test % bash comp-gene.sh \
-g /Users/gabriele/Rawdata/ShortSequences/genome1.fa /Users/gabriele/Rawdata/ShortSequences/genome2.fa \
-r /Users/gabriele/ReferenceFiles/MyReference.gbk \
-o /Users/gabriele/MyOutputFolder
```

Comparison of multiple genome assemblies, all in the same folder

```
-CompareGenomeV2.1test % bash comp-gene.sh \
-g /Users/gabriele/Rawdata/ShortSequences/genome*.fa \
-r /Users/gabriele/ReferenceFiles/MyReference.gbk \
-o /Users/gabriele/MyOutputFolder
```

CompàreGenome

User Guide to installation and use of **CompàreGenome**

For further information, please contact us!

Bioecopest srl - *Authorized GEP trials facility*

Technology Park of Sardinia - SP 55 - Loc. Tramariglio - 07041 Alghero - SS - Italy

VAT: IT02394290908 - www.bioecopest.com - info@bioecopest.com