Identification of Orthologous Genes and Pan-Genomics

Dr. rer. nat. Diego Mauricio Riaño-Pachón

Laboratório de Biologia Computacional, Evolutiva e de Sistemas

Centro de Energia Nuclear na Agricultura

Universidade de São Paulo

diego.riano@cena.usp.br

Analysing pan-genomes

Make sure there are not error messages after running each of the following commands!

Run the following command to install the required software for this tutorial:
```
sudo apt install gnuplot5 evince
```

Obtain the data for the Pan-genome exercises from Moodle, look for the file `genomes.pep.tar.gz,` download it to your computer, create a folder called pangenomics in your home, and move it there.

Download the software BGPA from https://sourceforge.net/projects/bpgatool/, unpack it, change permission so that the files muscle and BPGA-Version-1.3 have execution permission for all users

```
chmod a+x muscle BPGA-Version-1.3
export      PATH=~/Downloads/BPGA-1.3-linux-x86_64-0-0-0/BPGA-
Version-1.3/bin:$PATH
```

Download the software USEARCH from https://www.drive5.com/usearch/, make sure to rename the executable to usearch, and move it the same folder where you have the sequence files (~/pangenomics), make sure the usearch executable has execution rights. Copy also the usearch program inside the folder: `=~/Downloads/BPGA-1.3-linux-x86_64-0-0-0/BPGA-Version-1.3/bin,` move to that folder and run the BPGA-1.3 executable to generate the database files.

Move yourself to the folder ~/pangenomics, run the executable BPGA-1.3 and follow the on-screen instructions, inspect the output files and discuss with your instructor and classmates.

Identification of Orthologous genes

F0ollow the tutorial: http://lab.dessimoz.org/tutorials/omapractical/