Notes on RNA-Seq Tutorial, 5 June 2019

- We will be using single end, 100 bp reads. So whenever the tutorial mentions “paired-end reads” just ignore it.

Data:

- Upload all files from the https://github.com/dePamphilis/NSF-Summer-REU-Bioinformatics-Training/tree/master/data/rnaSeq_tutorial/ to your Galaxy history according to the methods outlined in the tutorial
- Combine all the “Control” files (S24C*) into a single collection and all of the “Pathogen” files (S24P*) into a single collection. This will make downstream steps much easier.
    - If you do not know how to do this, follow the instructions under “Creating collections in practice” from https://galaxyproject.org/tutorials/collections/

Mapping:

- I have already done the read mapping for you, which can take a long time. Use the files from “rnaSeq_tutorial_alignment” and “rnaSeq_tutorial_spliceJunction” to begin the featureCounts step in the tutorial

Visualization:

- We will skip visualization of aligned reads with IGV
    - Instead will look at reads mapped to the Criollo genome on the Criollo website
- Used ‘Heatmap with ggplot’ rather than heatmap2