libert-genomics-pipeline

Hello, stranger

this is a pipeline for running through all of the software necessary for genomic alignment

Contents

- `install.sh` – installs all necessary software and packages on a fresh computer or server
- `pipeline.sh` – wrapper for align.sh that puts everything in a tmux session
- `align.sh` – script that runs through the list of alignment software below

1. FastQC
2. cutadapt
3. FastQC (again)
4. Tophat
5. Cuffquant

How To Use

Installing

1. `ssh` into the server that you want to run the alignment on.
2. run the `install.sh` script if you don't have all the software you need yet.
3. if needed, `git clone` this repository somewhere on the server.
4. download and create all the `.gff`, `.bt2`, `.ver`, and `.fa` reference files that you need for the Dog genome and place them into the `dog` directory, same for `mouse` directory. Label them `latest_dog_genes`/`latest_mouse_genes`.
5. run `chmod +x *.sh` in this repo to make everything executable.

Commands

You are now ready to run the pipeline! First `cp` or `mv` the `.fastq` files you want into the same directory as the scripts.

```
./pipeline.sh start
```

- answer all the questions that it asks you, such as how many cores you want to use, the species you are aligning to, etc. Press `ENTER` through all the prompts to start the pipeline.

- the script makes a new `tmux` session and starts the `align.sh` script under it. If there is already a session open, it will run it under that session (I think).
- you can close out of the `ssh` session or terminal window that you are running from and the script will continue to run.

`./pipeline.sh attach`

- when you want to check on the progress of the pipeline, run this
- this will connect the `tmux` session to your current terminal or `ssh` window and show you the current output of the running pipeline

`tmux detach`

- you can run this inside the `tmux` window to get back out into your terminal or you can close the window but this is probably safer

.gitignore

the `.gitignore` file should ignore all the genomic files like

1. `*.bt2`
2. `*.fa`
3. `*.fa.fai`
4. `*.fa.tlst*.fastq`
5. `*.fastq.qz`
6. `*.gff`
7. `*.ver`

but if you see any git pushes taking an abnormally long amount of tmie, `Ctrl+C` that shit so we don't crash github.