TrimGalore on HPC

What is TrimGalore?

TrimGalore is a tool used for pre-processing high-throughput sequencing data generated from various platforms such as Illumina, SOLiD, and others. The software is specifically designed to remove low-quality reads, adapters, and contaminants from the reads. The tool works by trimming low-quality bases from the 3' end of the reads and removing adapter sequences, which can lead to biases in downstream analyses.

TrimGalore is a wrapper script that makes use of other open-source tools such as Cutadapt and FastQC to perform the trimming and quality control steps. The software can automatically detect the sequencing platform and adapter sequences used in the experiment and adjusts the trimming parameters accordingly. Additionally, it supports parallel processing, which makes it faster for large datasets.

Overall, TrimGalore is a widely used and effective tool for pre-processing sequencing data, and it can greatly improve the accuracy of downstream analyses such as genome assembly, variant calling, and gene expression analysis.

Links:

Official Website

Manual

Versions Available:

The following versions are available on the cluster:

- `TRIMGALORE 8.6.3`

How to load *TrimGalore*?

To load TRIMGALORE, use the following commands:

```
#Load the TRIMGALORE module
module load bio/trimgalore/0.4.2
```

To verify if the module is loaded correctly, use the following command,

```
# List all the module loaded in the environment
module list
```

How to use *TrimGalore*?

To use TrimGalore, you need to have your raw sequencing data files in a directory on your computer. Once you have your data files ready, you can open a terminal window and navigate to the directory containing your data files. Next, you can enter the command to run TrimGalore, specifying any desired options or parameters. For example, you can specify the minimum length for reads to be kept, the quality score threshold for trimming, and the type of adapter sequence used in your experiment. TrimGalore will then automatically perform the trimming and quality control steps on your data files, outputting the trimmed reads and quality control report to a specified directory. You can then use the trimmed reads for downstream analysis, such as genome assembly, variant calling, or gene expression analysis. TrimGalore is a straightforward and easy-to-use tool, and its comprehensive documentation makes it accessible to researchers with varying levels of bioinformatics expertise.

Here is a sample slurm script about how to use trimgalore,

```
#!/bin/bash
#SBATCH --job-name=trim_galore
#SBATCH -p main
#SBATCH --qos main
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --cpus-per-task=4
#SBATCH --time=01:00:00
#SBATCH --output=trim_galore_output.out
#SBATCH --error=trim_galore_error.err

# Load any necessary modules
module load bio/trimgalore/0.4.2

# Change to the directory where your raw sequencing data files are
located
cd /path/to/my/raw/data

# Run TrimGalore with multiple threads
trim_galore --paired --quality 20 --length 50 --cores 4 --output_dir
/path/to/my/trimmed/data R1.fastq.gz R2.fastq.gz

# End of script
```

Where to find help?

If you are confused or need help at any point, please contact OIT at the following address.

https://ua-app01.ua.edu/researchComputingPortal/public/oitHelp