VCFtools on HPC

What is *VCFtools*?

VCFtools is a software suite designed for working with variant call format (VCF) files, which are commonly used in genomic studies to represent genetic variation data. VCFtools provides a range of functionalities, including file manipulation, filtering, summarization, and visualization of VCF files. It can perform various tasks such as extracting subsets of data based on specific criteria, computing summary statistics, and generating plots. VCFtools is a powerful tool for analyzing and interpreting genetic variation data and has become a widely used software in the field of genomics.

Links:

Official Website

Manual

Versions Available:

The following versions are available on the cluster:

- `bio/vcftools/Mar1020`

How to load *VCFtools*?

To load VCFtools, use the following commands:

```
#Load the VCFTOOLS module
module bio/vcftools/Mar1020
```

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 *VCFtools*?

To use VCFtools, first, you need to have a VCF file to work with. Then, you can run the desired VCFtools command with the appropriate options and arguments to perform different operations on the VCF file. Some common VCFtools functions include filtering for specific variants or samples, calculating various statistics such as allele frequencies and linkage disequilibrium, and visualizing data with graphs and plots. The VCFtools manual provides detailed documentation and examples of how to use each command, which can be accessed through the command line by typing "vcftools --help".

```
#Look up manual
man vcftools
```

Here is an example of how to use vcftools,

```
#Download example files
svn checkout https://github.com/vcftools/vcftools/trunk/examples
```

More examples on following link.

Run vcftools on those files,

```
vcftools --vcf input_data.vcf --chr 1 --from-bp 1000000 --to-bp
2000000
```

Here is a slurm script demonstrating the usage,

```
#!/bin/bash
#SBATCH --job-name=vcftools
#SBATCH --output=vcftools.log
#SBATCH --time=1:00:00
#SBATCH --mem-per-cpu=2GB
#SBATCH --cpus-per-task=1
#SBATCH -p main
#SBATHC --qos main

# Load VCFtools module
module bio/vcftools/Mar1020

# Set input and output file paths
INPUT_FILE=/path/to/input/file.vcf.gz
OUTPUT_FILE=/path/to/output/file.vcf

# Run VCFtools command
vcftools --gzvcf $INPUT_FILE --remove-indels --recode --out
$OUTPUT_FILE
```

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