Pre-analyses: getting the data to make pretty plots

2024-09-03

Before running the code in R to make the pretty figures, there are some analyses that need to be done outside of R. This document details how to go about each of those analyses. Similar to the vcf-filtering.slurm file, there are multiple steps to this that are all in the same slurm script.

First, you will need to copy the slurm script for pre-analyses from the ecogen shared directory into your personal directory. To do so, run the following lines of code:

```
cp /mnt/home/ecogen/shared/scripts/pre-analyses.slurm /path/to/your/directory/scripts/pre-analyses.slurm
```

Once you have that copied into your scripts directory, open the script with vim editor and make sure to change all the path names and file names to your specific directories and file names.

Plink and admixture

Step 1: convert vcf to plink

Open the pre-analyses.slurm file you copied into your scripts directory with the vim editor. Remove the # from the lines under Step 1. The code should look like this:

```
# Step 1: convert vcf to plink
plink --vcf snps-filtered-4.recode.vcf --recode --out snps-filtered-plink
```

The rest of the file below step 1 should have # before each and every line. Once that is complete, you can run the code with the following command:

```
sbatch pre-analyses.slurm
```

This will output plink files – .ped and .map

Step 2: convert to binary plink

Open the pre-analyses.slurm file you copied into your scripts directory with the vim editor. Add in # for the previous step, and remove the # from the lines under Step 2. The code should look like this:

```
# Step 1: convert vcf to plink
#plink --vcf snps-filtered-4.recode.vcf --recode --out snps-filtered-plink

# Step 2: convert to binary plink
plink --file snps-filtered-plink --make-bed --out binary-snps-filtered
```

The rest of the file below step 2 should have # before each and every line. Once that is complete, you can run the code with the following command:

```
sbatch pre-analyses.slurm
```

This will output binary plink files – .bed, .bim, and .fam.

Step 3: perform admixture

Open the pre-analyses.slurm file you copied into your scripts directory with the vim editor. Add in # for the previous step, and remove the # from the lines under Step 3. The code should look like this:

```
# Step 2: convert to binary plink
#plink --file snps-filtered-plink --make-bed --out binary-snps-filtered

# Step 3: perform admixture
admixture binary-snps-filtered K
```

Replace the K with the number of clusters you want for the analysis. Once that is complete, you can run the code with the following command:

```
sbatch pre-analyses.slurm
```

This will output two admixture files ending in .Q and .P. .Q contains the estimated proportions of ancestry for each individual in the dataset, and .P contains the allele frequencies for each ancestral population.

Once completed, make sure to save all files created here to your computer.

currentNe

To estimate the effective population size, navigate into the currentNe directory that is on your premise space. Run the following code in there:

```
./currentNe /path/to/your/filtered/vcf/snps-filtered-4.recode.vcf #-of-chromosomes
```

This will run the currentNe effective population size estimator. Make sure to write in your vcf file path and insert the number of chromosomes where #-of-chromosomes is typed. The output file will have all of the Ne estimates. If you wish to run this for each subpopulation, do the same thing but only with a vcf that has your subpopulations of interest. Now you are ready to make pretty graphs!