Assignment 2

Reading

Nielsen et al. 2011 Genotype and SNP calling from next-generation sequencing data

McCormack et al. 2013 Applications of next-generation sequencing to phylogeography and phylogenetics

Homework

We will be using sequences available on google drive. You will need to download these onto your personal computer and use `scp` (or `PSCP.exe` if you're using putty) to move these onto your account on FARM.

1. **Align reads from one maize individual to the reference genome**

   Download the reference maize chromosome 10 from

   ftp://ftp.ensemblgenomes.org/pub/plants/release-30/fasta/zea_mays/dna//Zea_mays.AGPv3.30.dna_sm.chromosome.10.fa.gz

   Use the software bwa to align the `Zmays` sequences. To run much of the bioinformatics software on Farm, you will need to use the `module` command. For example, to make `bwa` available, use

   ```
   module load bwa/0.7.9a
   ```

   Feel free to explore the parameters of `bwa mem`, but for now we can just align using the default parameters. For information on how to run, load the module and type

   ```
   bwa
   ```

   Turn in your alignment script.

   Hints: Use `wget` to download the reference. Remember to index the reference with `bwa index`. Include relevant `module load` commands in your slurm script! Try using `srun` to log in interactively and practice commands until you have something that works. Then put that command in your script.

2. **Get summary statistics of the alignment**

Use samtools to get basic statistics of the alignment. Use `samtools flagstat`. What sequencing depth is your alignment?

Hints: Module avail can show you what software is on the cluster. Samtools is already there. `flagstat` needs a bam file input; is your alignment in bam format?

3. **Map two more genomes and get summary statistics.**

   Use your alignment script to now align reads from `Zmex` and `Tdip`. Run `samtools flagstat`, and answer the following questions with these alignments:

   - Why did so few reads map?
   - How are the summary stats different?
   - What might be an explanaton for any differences?

Challenge

Call SNPs! Use `samtools mpileup` to call SNPs from these samples. Do you get more or fewer SNPs than you expected? Any idea why?