P/BIO381 Ecological Genomics Spring 2020

Homework assignment #1 -- Due by Tuesday Feb 25, 2019 (end of day)

We've spent the past 4 coding sessions working with the red spruce exome data to gain proficiency with bioinformatics pipelines used in population genomics, and to begin gathering insight into the diversity present within range-edge populations of conservation concern.

Your assignment here is to summarize and distill your technical work on this module, and reflect on what you have learned about the populations genomics of this species.

Guidelines and expectations:

- The write-up should be 2 pages (max) single spaced, including tables/figures (references on a separate 3rd page)
- Approach the writing as a technical report based on the work you've done to date. That is, write for a scientific audience using appropriate technical language and narrative style, with citations used when referring to methods or making factual assertions.
- Your write-up should include the following essential elements:

    - Background (1-2 paragraphs):
        - A brief description providing context and motivation of the problem we're trying to address with these data
        - Brief background on the study species, biological samples, library prep, and sequencing strategy (look through early tutorials for info, plus your notes from class)

    - Bioinformatics Pipeline (2-3 paragraphs):
        - Detailed description of the various steps you used for the analysis of the sequencing data. Take it from QC assessment of the raw reads up to estimation of population genomic diversity.
        - This will section should demonstrate both your technical knowledge of the flow of the different steps in the pipeline, and your level of proficiency in understanding why each step was done. Include justification for using particular analysis approaches or choices as appropriate (e.g., Why did we map to a reduced ref instead of the entire *P. abies* reference genome?; Why did we use ANGSD instead of analyzing "hard called" genotypes?, etc...).

    - Results (1-2 paragraphs)
        - Report your findings from the different analysis steps. Use a combination of reporting results in-line in your text and summarizing more detailed information in tables and/or figures. You may use a max of 3 tables/figures

total. Be sure each table/figure has a title, and a very brief legend describing its contents.
■ Include in your results section both "methodological" results (numbers of reads, mapping stats, depth of coverage, etc) as well as "biological" results (estimates of the genetic diversity metrics, etc).

○ Conclusion (1-2 paragraphs)
■ Give your biological conclusion so far from the data: What have we learned about the diversity and demographic history in this set of populations? Relate back to your motivation given in the Background section.
■ Discuss any caveats or uncertainties that should be considered when interpreting the biological conclusions.
■ Discuss any methodological challenges encountered along the way that are relevant to your results and their interpretation.
■ Discuss opportunities for future directions.

○ References (listed on a separate page)
■ Cite papers in MLA format. Example:

Nystedt, Björn, et al. "The Norway spruce genome sequence and conifer genome evolution." *Nature* 497.7451 (2013): 579-584.

■ Only include references to papers you cite in your text

○ Github: Have your github lab notebook up to date, and any scripts used for your analysis available in your github "myscripts" folder.

You may collaborate with each other to discuss details and share notes, but your write-up should be done independently and represent your own work.

Due by Tuesday Feb 25, 2019 (end of day)

Email your homework in doc or pdf format to:

Stephen.Keller@uvm.edu
Melissa.Pespeni@uvm.edu
Lauren.Ashlock@uvm.edu