Title: Annotation of Olympia oyster (*Ostrea lurida*) and Pacific geoduck (*Panopea generosa*) genomes using WQ_MAKER on Jetstream cloud.

Primary Field of Science: Genomics
Resources Request Information:
· What percentage of the work you expect to do in this allocation will be the following types:
o Production (actually doing research): 100%
o Exploration/porting (preparing to do research): 0%
o Education (teaching others to do research): 0%
· What percentage of the jobs you expect to run in this allocation will be the following types:
o Submitted through command line/script: 100%
o Submitted using Grid tools (such as GRAM): 0%
o Submitted through a meta-scheduler: 0%
· Please estimate what percentage of the science runs you expect to perform in this allocation will be the following types:
o Independent (a job that is not immediately connected to any other job - a job that is artificially broken into chunks by queue limits should still be placed this category): 0%
o Independent but related (such as jobs that make up an ensemble or parameter sweeps): 0%
o Tightly-coupled (multiple jobs that will run simultaneously): 0%
o Dependent (multiple jobs such as in a workflow): 100%

Resource Requested: Jetstream
Resource Requested Amount: 100,000 SUs and 1TB disk space

Abstract:

MAKER is a popular genome annotation tool, but is relatively slow due to the way in which it it handles input scaffolds. Although MAKER can be configured to take advantage of high performance computing (HPC) clusters using mpi, it's difficult for most people to setup and run. An instance of MAKER (MAKER 2.31.9 with CCTools v3.1) was created on Jetstream by CyVerse in collaboration with Thain group at University of Notredome (Hazekamp et at., 2018) which not only alleviates the difficulties of trying to configure maker on an HPC, but also implements a Work-Queue (MASTER/WORKER) configuration which distributes the workload across multiple Jetstream instances, drastically increasing the scalability and performance/speed of MAKER and consequently genome annotation.

We have recently tested out the performance of WQ-MAKER using a CyVerse allocation on Jetstream in annotating a subset of the Olympia oyster genome using a limited number of WORKER on Jetstream instances and the increase in annotation speed is substantial, compared with testing on an HPC node running MAKER. The difference in time between running WQ-MAKER on Jetstream vs. running MAKER on the HPC node is on the order of months faster.

Our lab is requesting an allocation (100,000 SUs and 1 TB disk space) on Jetstream to annotate both the Olympia oyster and Pacific geoduck genomes. Both genomes are relatively large (> 1Gbp) and are likely to be complex (i.e. repetitive) making annotation more difficult (i.e. slower) than other eukaryotes. Both genomes have been *de novo* assembled from a variety of sequencing data (Illumina short read, PacBio long read, Chromium, and/or Hi-C). Annotation of these genomes will allow us to move forward with RNAseq, Methyl-seq, and proteomics projects that have been completed, but require annotated genomes to proceed. The Olympia oyster genome is ready to run immediately. The geoduck genome assembly will be finished by the end of August 2018 at the latest.

References

1. Hazekamp, N.L., Devisetty, U.K., Merchant, N., & Thain, D. (2018). MAKER as a Service: Moving HPC Applications to Jetstream Cloud. *2018 IEEE International Conference on Cloud Engineering (IC2E)*, 72-78.