MAFMC Quarterly Progress Report 5/5/2020

Surfclam Diagnostics & Population Connectivity project

M. Hare and H. Hartung, Cornell University

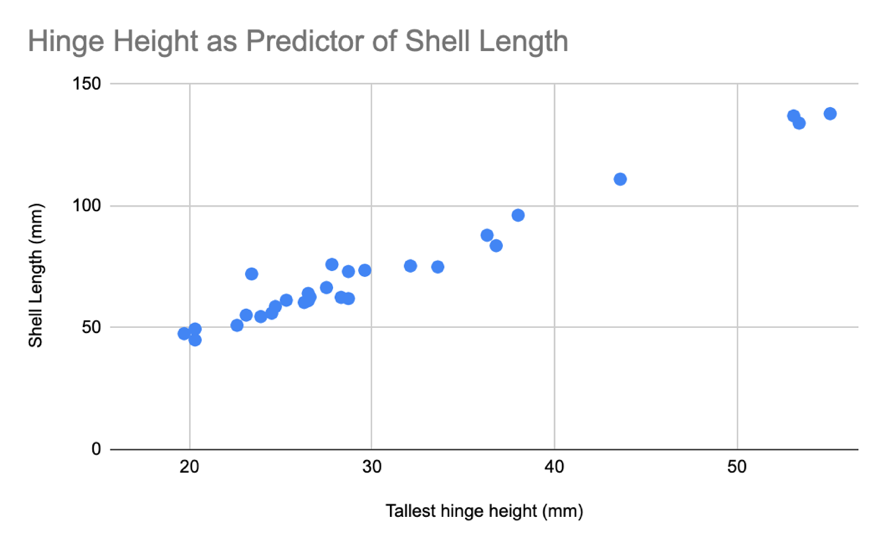
Activity this quarter was focused on analysis of *S. solidissima* and *S. similis* transcriptome sequences to search for multiple candidate DNA variants that can be useful for distinguishing these cryptic taxa and identifying hybrids. In addition, DNA quantification per sample and DNA quality control was conducted in preparation for a pilot run of genomic DNA sequencing. Late March campus shut downs halted progress on laboratory work and caused some disruption of the bioinformatic objectives.

Grant Objectives (*from proposal*):

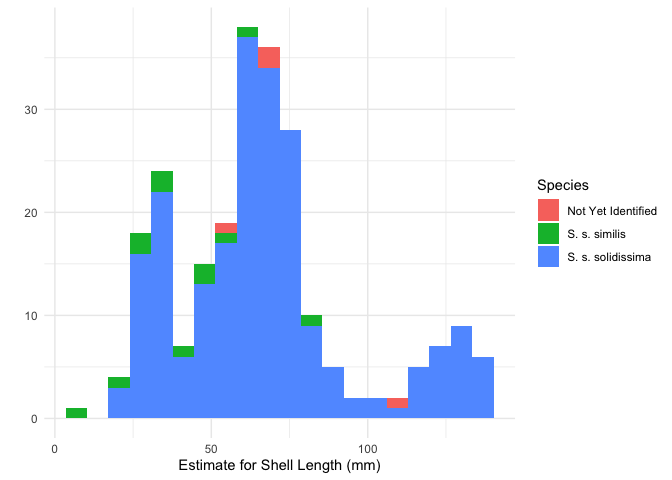
1. *Supply MA, NY, NJ and the federal survey team with sampling kits for proper preservation of tissue for DNA analysis during their surveys. Apply the species diagnostic to 3000 samples from nearshore survey sites where the two subspecies have overlapping range distributions. To the extent possible, collect and analyze samples in such a way that depth can be tested as a habitat variable with differential subspecies affinities.*

Samples are in hand from around Long Island, NY, and from Georges Bank (Federal survey). Efforts continue to get nearshore samples from Massachusetts and New Jersey where state governments have no scheduled survey intentions. In NJ we have been in contact with Kira Dacanay at NJ Bureau of Shellfisheries and got coordinates for possible *S. similis* populations. It may be possible to get James Nickels at Monmouth University to sample some sites in conjunction with classes he teaches (once instruction resumes). Federal survey sampling from Nantucket Shoals and offshore New Jersey was scheduled for this summer but could be delayed.

2. *Because New York indicated an inability to sample outside their standard survey design, contract with a fisherman to do targeted sampling around Long Island, NY.*

Here we report size data on our Long Island, NY samples. Out of 238 measurable samples, 40 specimens had incomplete broken shells, but the hinge was intact. We tested whether shell hinge dimensions could be used to approximate shell length (figure at right). The strong correlation between maximum hinge tooth height and shell length confirms that hinge dimensions will provide useful clam size predictors.

Additionally, a frequency histogram for all measurable Long Island samples (below) is shown for *S. solidissima* and *S. similis* based on a mitochondrial DNA diagnostic. A multi-modal distribution for the *S. solidissima* data is consistent with several distinct cohorts along the South shore of Long Island, NY. Maximum shell length for *S. similis* was 83mm and for *S. solidissima* was 140 mm.



3. *Develop a species diagnostic assay based on three nuclear DNA markers that can be applied at low cost to identify first generation hybrids as well as subspecies.*

In progress, but nothing new to report.

4. *Extract total RNA (gene transcripts) from two samples of each taxon and generate sequence data for the full transcriptome of expressed genes. Assemble these sequences de novo into a transcriptome “reference” for each subspecies for use in whole genome sequence analysis and to design a species diagnostic.*

Transcriptome assembly completed and being used as the reference for variant calling among transcriptome samples.

5. *Collect whole genome sequence data from 350 samples (175* S.s. similis*, including an existing n=25 Hare Lab sample from Georgia, 150* S.s. solidissima*, and 25 likely hybrids) and identify DNA variants within and between each subspecies.*

Pilot genomic libraries for 4 *S. similis* and 4 *S. solidissima* samples have been prepared and await sequencing when the facility opens up again. Pilot data are important for optimizing the molecular details of genomic library prep to make sure we end up with a sufficient number of well-sequenced loci for our connectivity analysis.

6. *Analyze and report on population connectivity among populations within each taxon using methods that establish the geographic scale of gene flow and evolutionary independence. Genome scans will be used to identify nonneutral loci (candidates potentially experiencing directional selection) and separate analyses will estimate population subdivision in neutral and nonneutral classes of loci.*

Nothing new to report.