Mid-Atlantic Fishery Management Council
800 North State Street, Suite 201, Dover, DE 19901
Phone: 302-674-2331 | FAX: 302-674-5399 | www.mafmc.org
Michael P. Luisi, Chairman | P. Weston Townsend, Vice Chairman
Christopher M. Moore, Ph.D., Executive Director

MEMORANDUM

Date: March 23, 2022

To: Council

From: Jessica Coakley, Staff

Subject: Presentation on research project entitled, "Surfclam species diagnostics and population connectivity estimates to inform management"

The Council contracted researchers at Cornell University to examine species connectivity among the commercially important Atlantic surfclam, *Spisula solidissima solidissima*, and its sister-taxon the Southern surfclam, *Spisula solidissima similis*. While these taxa are impossible to distinguish in the field, they are easily distinguished using genetic markers. Atlantic surflcam population structure and connectivity are important factors that shape the types of management approaches needed to maintain sustainable surfclam harvests.

Despite some delays and sampling challenges associated with COVID-19, the project is nearing completion. At the April Meeting, Dr. Mathew Hare and Hannah Hartung will present the Council with the results of this project to date. A summary of this work has been provided behind this tab.

The final report from the project is anticipated in June 2022, with a presentation on those results to be given to the Scientific and Statistical Committee at their July 2022 meeting.

Preliminary summary of results on surfclam population structure and population connectivity
Matt Hare & Hannah Hartung, Cornell University
mph75@cornell.edu
Federal Identification Number: 51-6148342

Key Project Objectives:

- Generate sequence data for the full transcriptome of expressed genes in both subspecies. 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.
- 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.
- 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.
- 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.
- Collect genome-scale data from 350 samples and identify DNA variants within and between each subspecies.
- Analyze and report on population connectivity among populations within each taxon using methods that establish the geographic scale of gene flow and evolutionary independence.

Background - the utility of genomics

Scientists focus on differences. In biology, whether the goal is taxonomy, molecular biology, or fisheries management, one of the greatest challenges is the need to evaluate how meaningful differences are for the goals at hand. In some cases, when advances provide higher resolution discrimination of differences, like our recent ability to detect transcription from 70-80 percent of the human genome (protein coding genes account for only 1-2% of the genome), it opens up discoveries that lead to whole new definitions of "function" vs "junk", relevant signal vs noise.

Genomic-scale assays of DNA variation, and the ability to apply these to population samples from taxa with no genomic resources (i.e., no reference genome sequence), are enriching the longstanding contributions of genetics to population studies. However, it typically is not obvious what demographic and population biology meaning to place on subtle population genetic differences. Interpretation requires careful consideration of population processes happening at both ecological and evolutionary time scales, and molecular impacts of those processes on functional DNA variation influencing relative fitness (selection, gene flow and genetic drift) as well as on "neutral" variation not affecting fitness (gene flow and genetic drift). Unlike demographic and ecological studies of population biology, where the boundaries of a distinct population often are defined by the impact of immigration on population growth, or the degree of independence between vital rates, population genetic differentiation is most informative about reproductive interactions (random mating within populations, gene flow among populations). Depending on the study design and context, population genetic variation can be

used to estimate contemporary processes (using high resolution multilocus genotypes as 'tags' for tracking movements in recent generations), or to infer average processes in the recent evolutionary past, typically with the benefit of an evolutionary model.

Cryptic surfclam subspecies are partially sympatric 'good' species

Taxa that have evolved some measure of reproductive isolation but are still phenotypically the same where they co-occur, are ideal subjects for informative genetic analysis. For a long time the two nominal subspecies of surfclam, *Spisula solidissima solidissima* and *S.s. similis* were thought to be largely allopatric, with the latter rarely occurring north of Cape Hatteras, if at all, and confined to nearshore waters. Thus, observations of life history differences between inshore vs. offshore populations of *S.s. solidissima* have been interpreted solely as the plastic phenotypic consequences of inshore/offshore environmental differences or density effects (Jones et al. 1978; Ropes 1979; Jones 1980; Ambrose et al. 1980; Cerrato and Keith 1992).

Hare and Weinberg (2005) and then Hare et al. (2010) used genetic markers to demonstrate the presence of *S. solidissima similis* in Southern New England, including the previously fished surfclam population in Long Island Sound, NY. The reported genetic patterns were interpreted as consistent with full species status because the degree of genetic differentiation would be unlikely if gene flow were continuing between these two taxa, and sampling showed co-occurence of these taxa in Southern New England. However, shell morphometric analysis did not yield any traits or combinations of traits that easily distinguish these taxa.

High resolution genomics reveals an additional cryptic taxon

Sampling of *S.s. solidissima* for this project was hampered by the pandemic. We acquired Georges Bank samples from the federal survey but Nantucket Shoals and Delmarva shelf samples were obtained from commercial sources. Additional federal samples from 1999 and archived by M. Hare also were analyzed. Closer to shore, samples included a 2012 sample from the New York State DEC survey along the South shore of Long Island, and 2019 samples from the same region collected in shallow water near inlets by a contractor for this project. Additional *S.s. solidissima* samples were obtained from Massachusetts state surveys in Cape Cod Bay and south of Cape Cod. Samples of *S.s. similis* were based on effort by the contractor along the North shore of Long Island, Massachusetts state survey efforts, and a 2012 contract to sample the Georgia population (previous federal Hatch funds). The location of all samples is shown in Fig. 1. Note that mixed populations of these two taxa occur only south of Cape Cod. Only *S.s. similis* was found in Peconic Bay (end of Long Island) and in Long Island Sound, and only *S.s. solidissima* was observed along the South shore of Long Island.

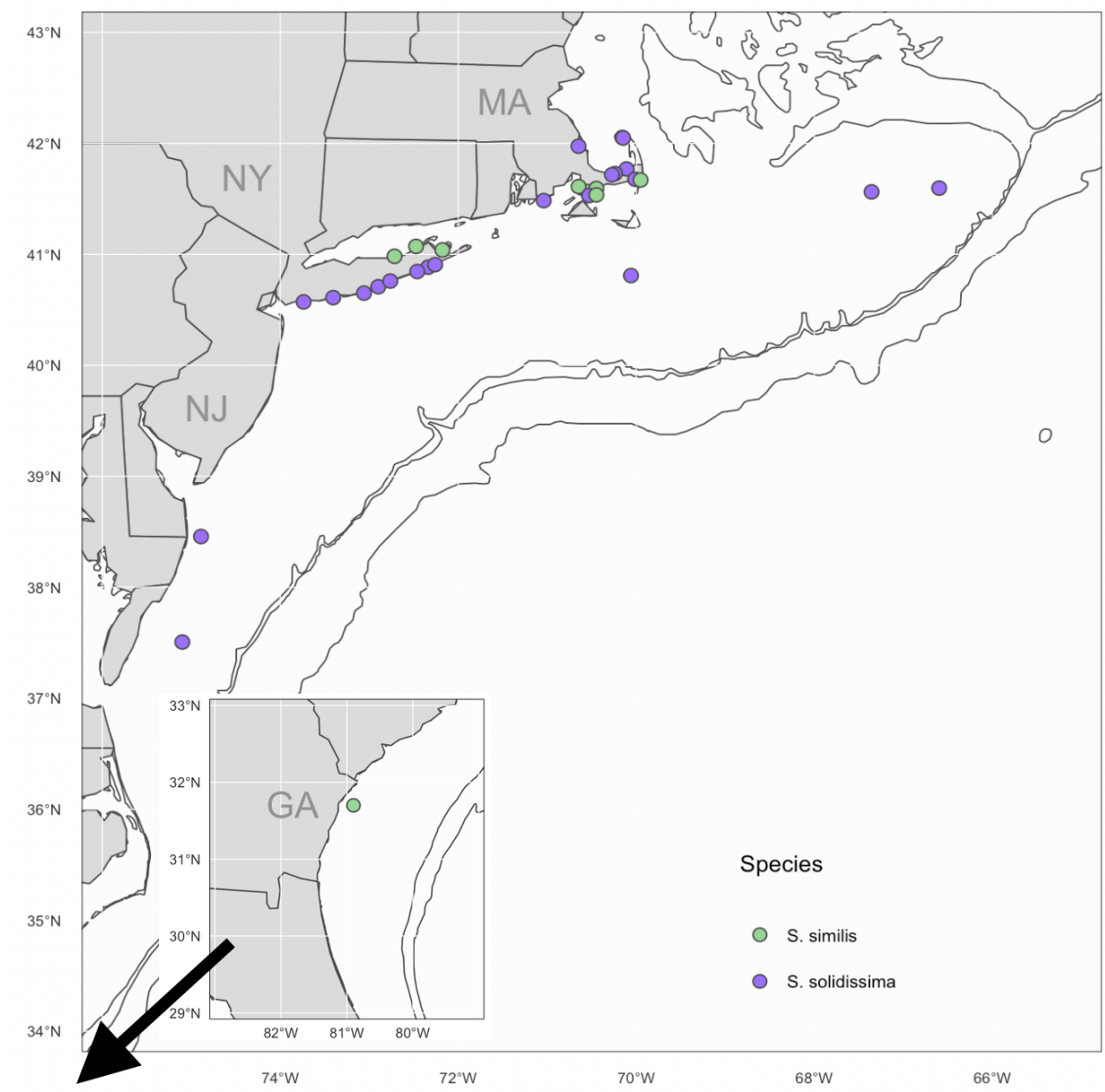

Fig. 1: Sample collection locations for *S.s. solidissima* (purple) and *S.s. similis* (green) used in genomic analyses.

Using a 'reduced representation' method of randomly sampling surfclam genomes, so that the same homologous chromosomal positions are sampled in each individual, we now have a high resolution dataset consisting of 2.6 thousand quality-filtered single nucleotide polymorphisms (SNPs) from chromosomal loci scattered through the genome. This dataset consists of loci that have been carefully selected to be comparable (i.e. homologous) between *S.s. solidissima* and *S.s. similis*. Larger numbers of high quality loci and SNPs have been identified for analyses within each taxon.

Using principal component analysis to explore the multidimensional allele frequency variance across individual specimens in both nominal taxa, the greatest variance explained by PC1 and separates *S.s. solidissima* from *S.s. similis* (Fig. 2). Surprisingly, *S.s. solidissima* samples show extensive allele frequency variance along PC2. The allele frequency variance explained by these two PC axes, 9.18% and 7.99% respectively, indicate that the two *S.s. solidissima* clusters have nearly as much allele frequency differentiation between them as found between the two nominal subspecies. For now, we are referring to these two clusters as Genotype A and Genotype B.

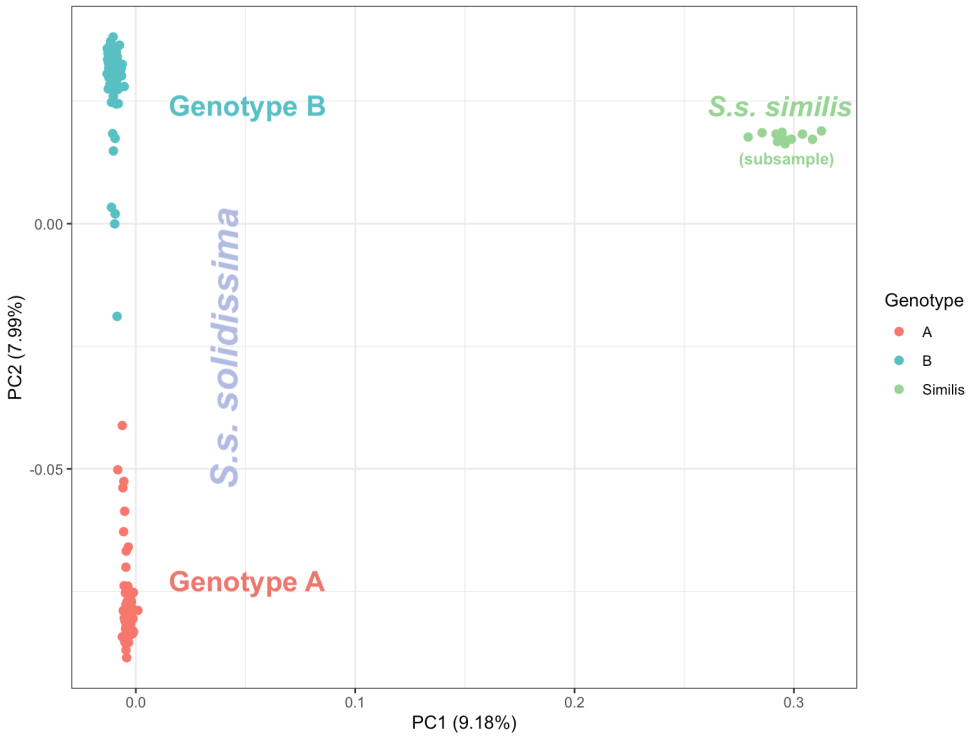

Fig. 2: Principal components analysis plot of PC1 and PC2 summarizing allele frequency differentiation among individuals from both nominal subspecies. Genetically differentiated clusters of S.s. solidissima are labeled Genotype A and Genotype B.

Spisula s. similis

Using PCA to explore patterns of population differentiation among all samples of *S.s. similis*, using 12.7 thousand SNPs, the three geographically discrete groups of samples show genetic differentiation (Fig. 3). The greatest differentiation along PC1 (2.88% allele frequency variance explained) separates Southern New England (NY+MA) from Georgia. Along PC2 the differentiation between samples from the North shore of eastern Long Island and Peconic Bay (NLI) versus the Southern coast of Massachusetts (SCC) is subtle, but it is interesting that there is any distinction at all. Using F_{ST} as a metric of allele frequency differentiation that spans from 0 to 1.0, the latitudinal contrast has average F_{ST}=xx whereas the two southern New England populations have average F_{ST}=xx. Our ongoing analyses are testing hypotheses about demographic history (e.g. historical population bottlenecks or admixture) that might explain the distinct patterns in NY and MA populations.

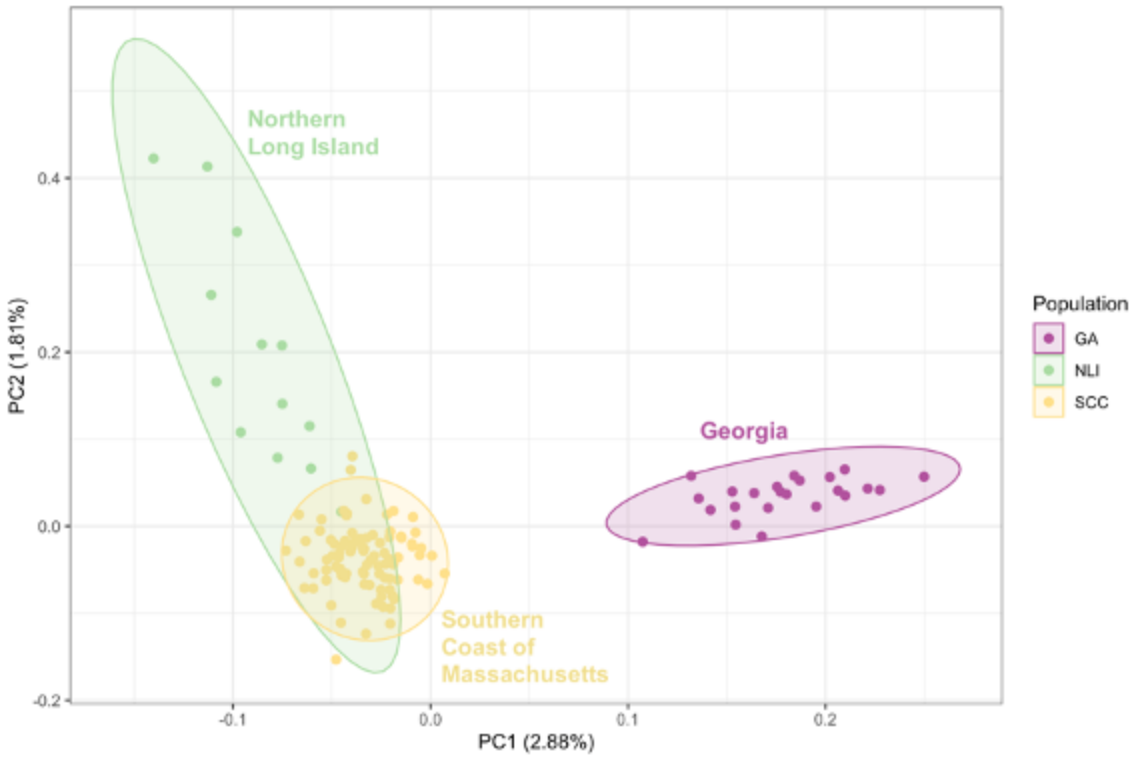

Fig. 3: Principal component analysis of all *S.s. similis* samples based on 12.7 thousand SNPs.

Spisula s. Solidissima

Focusing analysis on *S.s. solidissima* yielded 49 thousand high quality SNPs. Genotype clusters A and B are largely partitioned along PC1, whereas PC2 shows some distinction within Genotype B between clams from Cape Cod Bay and the rest of the Genotype B cluster (Fig. 4). Some Cape Cod Bay clams also show intergradation between the A and B genotype clusters. To examine this in more detail we used a model-based analysis that infers how many differentiated source populations are contributing to the observed genotypic variation, and at the individual level, whether genomic variation is best explained by a hypothesis of admixture (interbreeding between the hypothesized source populations). Admixture inferred with this model is more likely to be recent, not ancient. One way to think about admixture is with expectations from a pedigree when starting with two genetically distinct parents - the first generation offspring will have 50/50 genomes consisting of homologous paternal and maternal chromosomes. If F1 individuals backcross to a parental type, the expected proportionality in the F2 generation is 75/25, and so forth. The history of interbreeding is likely to be complicated and there are many histories that could produce a 75/25 pattern in an individual, but in general between to very distinct source populations, moderate admixture (50/50) is likely to be more recent and minor admixture (90/10) is older.

As with the PCA, the greatest number of admixed individuals was found in Cape Cod Bay, but only a small minority of clams (15%) had moderate levels of admixture (Fig. 5). The admixed clams were scattered all around Cape Cod Bay and no ‘pure’ type A clams were found. New Jersey had the same pattern, but with only slight admixture in only a few specimens. Southern Cape Cod and Southern Long Island were the only regions where ‘pure’ genotype A

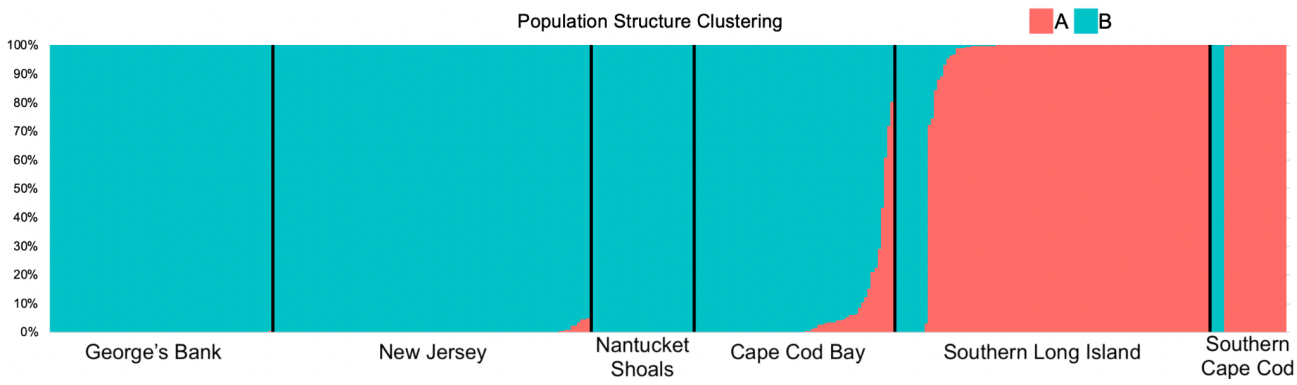

Fig. 5: Clustering and admixture results from program STRUCTURE applied to all S.s. solidissima samples using 2.6 thousand SNPs. Models assuming K=2 source populations showed the greatest support from the data, here depicted as blue-green for the Genotype B source population and orange for the Genotype A source population. Black vertical lines separate population samples. Each individual specimen is represented with a thin bar that is either blue-green, orange or a combination indicating proportional contributions from these two sources (admixture). In mixed populations, individual clams are ordered from fully type B to increasing proportions of type A.

and B co-occurred, meaning they had the opportunity for interbreeding. Southern Long Island also included a few admixed clams.

Patterns of A/B admixture and sample sizes are shown on a map in Fig. 6 for Southern New England. We are making efforts to compare the depth distribution of genotype A vs. genotype B clams, and to analyze length by age patterns for the subset of genotyped clams that were aged (by federal and NY state labs). In both cases a very uneven distribution of samples makes interpretation difficult. Shells are being sent to the Woods Hole NOAA lab for aging to improve our ability to estimate von Bertalanffy growth curve parameters for both genotype A and genotype B populations.

Genetic Diversity

To our surprise, all surfclam populations have similar levels of genetic diversity as measured by one of the most sensitive indicators, allelic richness (i.e., the average number of alleles per locus in a population after correcting for sample size differences). Typically a SNP locus only has two alternate nucleotides segregating in the population. Instead, for allelic richness we analyzed nearby SNPs jointly as a haplotype, so for a haplotype consisting of 3 SNPs we might distinguish alleles AGG, AGT, TGG, TGT, ACG, ACT, TCG, TCT. Structuring the SNP data this way provides a measure of genetic diversity that is more sensitive to recent fluctuations in population size. For example, this single locus haplotype example might show 8 alleles in a large population but only 5 alleles in a numerically small population. Allelic richness averaged near 3 for all three S.s. similis regional populations and for S.s. solidissima genotype A and genotype B populations.

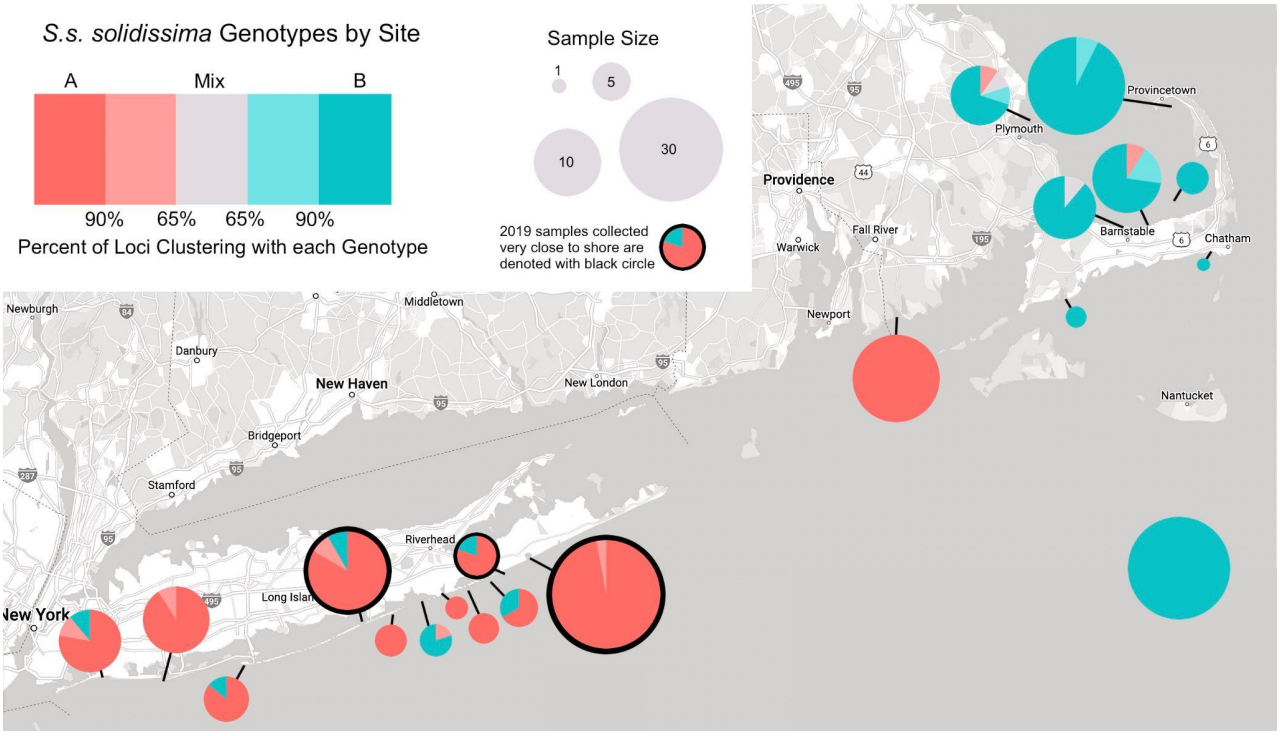

Fig. 6: Southern New England pie diagrams depicting sample size and distribution of genotype A, genotype B and admixed individuals.

Model based estimates of genetically effective population size and demographic history are ongoing. For now, here is our informal suggestion for why a sparse nearshore population of *S.s. similis* might have comparable levels of genetic diversity compared to an abundant, commercially important taxon distributed across much more extensive habitat. If contemporary demographic processes are limiting *S.s. solidissima* effective population size (genetic diversity) relative to *S.s. similis* then the most likely candidates are a relatively higher variance in reproductive success or a more skewed sex ratio in the former, both factors that lower effective population size below census size. Alternatively, because historical fluctuations in population size can reduce genetic diversity and lower effective population size, *S.s. solidissima* may have experienced greater Pleistocene or post-Pleistocene population size fluctuations than *S.s. similis*. The third possibility for metapopulations (both taxa include an array of regional populations connected by larval dispersal) is that *S.s. similis* experiences greater large scale gene flow relative to *S.s. solidissima*, effectively enlarging its genetically effective population size. Analyses of gene flow are ongoing, and are challenging to compare between taxa for a test of this hypothesis.

Gene Flow

The preliminary result on gene flow that can be shared at this point is based on a population genetic estimation of relative gene flow magnitude and directionality. The program divMigrate uses a novel pairwise population analysis approach to test whether the differences in

allele frequencies, and the pattern of private vs shared alleles between two populations, supports a model of asymmetric gene flow. The populations are assumed to be at an evolutionary equilibrium between the homogenizing force of gene flow and differentiation caused by genetic drift. Every pairwise comparison is made between subpopulations to determine which pairs have statistically supported nonzero gene flow, and the direction of exchange if the gene flow is significantly asymmetric. The report of results from a biophysical model by Zhang et al. (2016) suggested that *S.s. solidissima* larvae are strongly advected to the southwest, like a conveyor belt from Georges Bank to the Delmarva shelf. The genomic results we report here are the first opportunity we are aware of to empirically address their model based predictions.

In *S.s. similis* the highest relative level of gene flow was inferred to be between the Massachusetts and Georgia populations with a strongly southwestern directionality. The connection between Massachusetts and Long Island Sound surfclams also was directional and westward, with only half the magnitude of gene flow relative to MA->GA. This result is based on 1250 haplotype loci, using 10,000 bootstraps for the statistical significance testing.

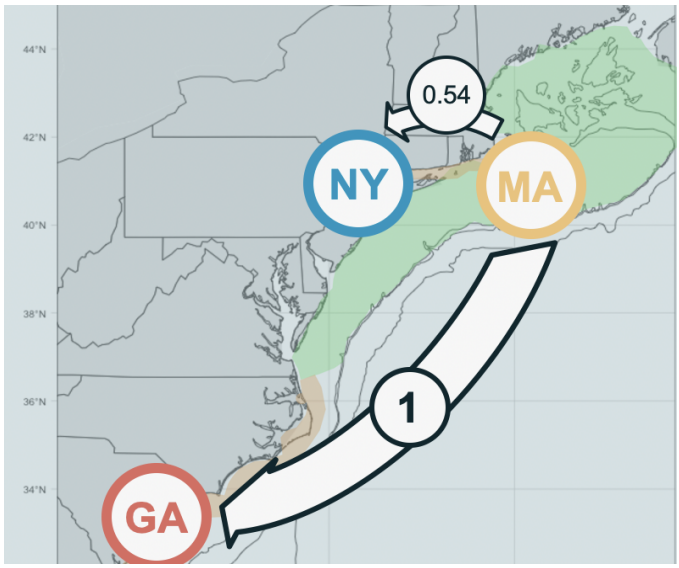

Fig. 7: divMigrate inference of gene flow directionality
and relative strength. The maximum observed gene
Flow is labeled “1” and other gene flow levels are relative
To that.

Gene flow inference with S.s. solidissima is ongoing. Preliminary results will be included in the presentation.

Summary and Conclusions

We want to emphasize that this summary describes the preliminary results from many gigabytes of genomic data obtained only a couple months ago. We are presenting patterns that seem robust and reliable, but many checks and further analyses are still in the works.

The most dramatic new information is population subdivision within *S.s. solidissima*. Patterns in the data suggest that the genomic differentiation between genotype A and B is not driven by a small number of markers showing extreme differences. Instead, population subdivision in both *S.s. solidissima* and *S.s. similis* seems to involve small differences at many loci such that the cumulative signal from genomic scale sampling was necessary to detect differences. This pattern of differentiation, consistent with slow genetic drift between large populations, may make it difficult or impossible to find one or a few diagnostic loci for easy genetic screening of samples. Determining the minimum effort required (loci to genotype) to discriminate the nominal subspecies or the two types of *S.s. solidissima* is one of our analytical goals.

Literature Cited

Ambrose WG, Jones DS, Thompson I (1980) Distance from shore and growth rate of the suspension feeding bivalve, *Spisula solidissima*. Proc Natl Shellfish Assoc 70:207–214

Cerrato RM, Keith DL (1992) Age structure, growth, and morphometric variations in the Atlantic surf clam *Spisula solidissima*, from estuarine and inshore waters. Mar Biol 114:581–593

Hare, M.P. and Weinberg, J.R., 2005. Phylogeography of surfclams, *Spisula solidissima*, in the western North Atlantic based on mitochondrial and nuclear DNA sequences. Marine Biology, 146(4), pp.707-716.

Hare, M.P., Weinberg, J., Peterfalvy, O. and Davidson, M., 2010. The “southern” surfclam (*Spisula solidissima similis*) found north of its reported range: a commercially harvested population in Long Island Sound, New York. Journal of Shellfish Research, 29(4), pp.799-807.

Jones, D. S., Thompson, I., Ambrose, W. (1978). Age and growth rate determinations for the Atlantic surf clam *Spisula solidissima* (Bivalvia: Mactracea), based on internal growth lines in shell cross-sections. Mar. Biol. 47:63-70

Jones DS (1980) Annual cycle of shell growth increment formation in two continental shelf bivalves and its paleoecologic significance. Paleobiology 6:331–340

Ropes JW (1980) Biological and fisheries data on surf clam, *Spisula solidissima* (Dillwyn, 1817). US DOC, tech. ser. rep. no. 24, Northeast Fisheries Science Center, Woods Hole, Mass., USA

Zhang, X., Munroe, D., Haidvogel, D. and Powell, E.N., 2016. Atlantic surfclam connectivity within the Middle Atlantic Bight: mechanisms underlying variation in larval transport and settlement. *Estuarine, Coastal and Shelf Science*, *173*, pp.65-78.