Surfclam Cryptic Taxa, Population Structure and Population Connectivity

Matt Hare & Hannah Hartung, Cornell University
mph75@cornell.edu
Federal Identification Number: 51-6148342
June 2022

Contents

Executive Summary

Management of surfclams has supported an economically important fishery and maintained strong stocks in most habitats. However, several unknowns (assumptions) create possible vulnerabilities, or opportunities whereby new knowledge can help further optimize management procedures. Ambiguities include the taxonomic level at which known taxa should be regulated (species or subspecies); taxon range distributions and habitat limits; and relative levels of demographic and evolutionary connectivity among populations. This project used genomic analyses to address and clarify each of these assumptions.

Under current classification, the two surfclam taxa inhabiting the western North Atlantic include *Spisula solidissima solidissima* and *Spisula solidissima similis* subspecies. They were previously described as having minimally overlapping range distributions North and South of Cape Hatteras, respectively, but recent work has demonstrated a robust population of *S.s. similis* in Southern New England (Hare and Weinberg 2005; Hare et al. 2010). Nearshore dredge sampling conducted for this project documented *S.s. similis* on the North shore of Long Island,

in Peconic Bay at the end of Long Island, but NOT along the Long Island south shore (sampling in inlets but not lagoons). No *S.s. similis* were found in Cape Cod Bay, suggesting that the southern coast of Cape Cod is their northern limit. Results reported here suggest habitat affinities for *S.s. similis* in Southern New England are similar to what has been described for this taxon in Georgia (Walker, 1998); shallow nearshore waters with swift tidal currents. A previous study used nuclear DNA gene trees to argue that these two nominal subspecies, *S.s. solidissima* and *S.s. similis*, are fully independent species because genealogical variation from each subspecies formed a monophyletic clade in each of two genes (Hare and Weinberg 2005). However, few individuals were analyzed and two genes represent a small sample from genomes that may have a mixture of patterns (as expected for recent species). In this report the case for full species-level classification is based on genealogical patterns at hundreds of genes. Perhaps even more persuasive, and more directly addressing a major tenet of the biological species concept, is the fact that no hybrids between these nominal subspecies were found in Massachusetts where both subspecies occurred in dredge and diver collections. Together these lines of evidence strongly support species level classification of *S.s. solidissima* and *S.s. similis*.

The assumption that *S.s. solidissima* represents one demographic unit, with strong connectivity between partitions such as Georges Bank and the Delmarva shelf, is only partly confirmed by results presented here. Population genomic analyses using several modeling frameworks all consistently indicated moderate gene flow connectivity among the continental shelf populations from Georges Bank to Delmarva and including Nantucket Shoals. Although a previously published biophysical model predicted unidirectional larval dispersal toward the Southwest among these populations, the genomic data provide only a very slight hint of this asymmetry.

A striking departure from genetic homogeneity was found when comparing offshore vs. inshore *S.s. solidissima* samples. Except for a few hybrid individuals, *S.s. solidissima* contains two genomically divergent populations that we currently are referring to as operational taxonomic units (OTU) A and B. These two *S.s. solidissima* OTUs are 2-3 times more genealogically distinct as the Georgia and Southern New England populations of *S.s. similis* are to each other. Genomic patterns and the occurrence of hybrids are consistent with a subspecies-level distinction between *S.s. solidissima* OTUs A and B. Nonadmixed OTU A individuals were exclusively found inshore and south of Cape Cod. Samples from federal waters were all OTU B, but OTU B also occurs at low frequency in Southern New England and at high frequency in Cape Cod Bay. Mixed populations of OTU A & B surfclams occurred along the South shore of Long Island and in southern Massachusetts, but hybrids were only in the former region. Hybrids between OTUs A and B of *S.s. solidissima* also were found in Cape Cod Bay, but these appear to be the result of OTU A immigration and interbreeding because no pure A type individuals were found. In terms of population structure and gene flow within these OTUs, the A type populations in Southern New England have genomic patterns consistent with slightly more East - West differentiation than expected under isolation by distance. The B clade shows more differentiation inshore vs offshore than within either of those regions, with the strongest gene flow barrier isolating Cape Cod Bay.

In summary, results presented here support consideration of the following modifications to regulatory and management assumptions:

1. Treat *S.s. solidissima* and *S.s. similis* as full species under the biological species concept.

2. Maintain continental shelf populations (OTU B of *S.s. solidissima*) as one management unit.
3. Manage Cape Cod Bay (OTU B of *S.s. solidissima*) as a demographically separate population and investigate the source of introgression from OTU A.
4. Manage inshore *S.s. solidissima* populations as a distinct subspecies (OTU A, but mixed with some B type) and investigate the depth distributions and life history differences between OTU A and B surfclams.

Background

Cryptic Surfclam Nominal Subspecies are Full Species

Population genetic analysis can be especially informative with taxa that have evolved some measure of reproductive isolation but are still phenotypically indistinguishable where they co-occur. For a long time the two nominal subspecies of surfclam, *Spisula solidissima solidissima* and *Spisula solidissima 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). Also, the biological species concept is difficult to apply to allopatric populations because one of its important criteria is the propensity to interbreed and produce reproductive offspring.

Hare and Weinberg (2005) and then Hare et al. (2010) used genetic markers to demonstrate the presence of *S.s. 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 genealogical pattern of differentiation at several genes (mitochondrial and nuclear) would be unlikely if gene flow were continuing between these two taxa, and sampling showed co-occurence of these taxa in some Southern New England localities. However, shell morphometric analysis did not yield any traits or combinations of traits that easily distinguish these taxa (Hare and Weinberg 2005).

The Utility of Genomics to Address Demographic Management Questions

It typically is not obvious what demographic and population biology meaning to place on subtle population genetic differences. Demographic and ecological studies define distinct populations in terms of the impact of immigration on population growth, or the degree of independence between vital rates. In contrast, population genetic differentiation is most informative about reproductive interactions. For example, genetics can reveal deviations from random mating within populations, and allele frequency differences across space are indicative of either limited dispersal connectivity or limited reproduction by immigrants. In addition to focusing on different population distinctness criteria, demographic measures typically apply to contemporary populations whereas population genetic patterns are largely shaped over the recent evolutionary past. Thus, evolutionary theory is used to analytically infer gene flow, divergence and admixture processes, then these can be translated into contemporary levels of relative demographic connectivity under the assumption that factors shaping population

exchange in the recent past (100s of generations) are representative of contemporary forcing factors. This may seem ill advised in a dynamic, nonequilibrium world. However, in a dynamic environment, demographic analysis of contemporary populations also provides insights with low relevance to future conditions, and if study duration is short it is possible to sample in a nonrepresentative year and be misled. The temporally shallow evolutionary approaches provided by population genetics are a cost-efficient approach to measure the population consequences of average demographic and evolutionary processes. A population genetic snapshot is less sensitive to episodic demographic events.

Population genetics requires two sampling processes - individuals from populations and loci (markers) from the genome of individuals - to observe patterns of genetic diversity that can be interpreted in terms of evolutionary and demographic processes. Now that genomics enables sampling thousands of variable markers as opposed to a dozen microsatellite markers, the increased inferential power is particularly important for resolving subtle differentiation among large populations with gene flow, as well as distinguishing loci influenced by natural selection versus those with negligible effects on fitness. It is the latter class of markers, sometimes referred to as selectively neutral, that we are focused on here for population connectivity inferences. Neutral genetic variants change in population frequency by genetic drift, and there is a deep body of theory available to interpret spatial patterns of differentiation in terms of gene flow connectivity and speciation (the evolution of reproductive isolating barriers).

Depending on the study design and context, population genetic variation can be used to estimate recent evolutionary processes by applying an evolutionary model. For example, observed population differences are compared to that expected from spatial models in which the differentiating force of genetic drift becomes balanced by the homogenizing force of average gene flow. The equilibrium pattern of spatial differentiation is referred to as isolation by distance and is reflected in a positive correlation between geographic distance and genetic differentiation (because for a given average dispersal distance, more proximate populations will exchange more migrants than geographically distant populations). Genetic drift changes allele frequencies more rapidly in smaller populations. Thus, all else being equal, large populations drift apart in allele frequencies relatively slowly compared with small populations. The time to isolation by distance (IBD) equilibrium, roughly speaking, is faster under high gene flow and with smaller populations. At levels of gene flow typical in marine systems (proportion of immigrants, m, =0.01 to 0.1) the time to equilibrium levels of differentiation is likely on the order of tens of generations, and some population subdivision can be expected much sooner (at 50% to equilibrium; Hardy et al. 1999). Thus, some population structure can be expected across geographically distant populations even when high gene flow (m) maintains stepping stone connectivity, and even if processes leading to the differentiation are relatively recent. With isolation by distance as a null model, spatially discrete deviations from IBD expectations can indicate relatively more or less gene flow connectivity than expected at a given geographic scale, for example if there is a seascape barrier to gene flow or a connectivity conduit.

Levels of gene flow connectivity are not necessarily equivalent to demographic connectivity. Here, to extend evolutionary connectivity inferences to the demographic exchange more relevant for population management, we will assume that across most of the studied

surfclam range the fitness of settling larvae is uniform relative to larval source population. In other words, not local adaptation among the studied populations.

Results and Discussion

Surfclam Sampling

A total of 541 *S.s. solidissima* were collected for this project, both from offshore federal waters and inshore NY and MA areas. Combined with archived samples the total *S.s. solidissima* sample size was 920. For this project a contractor did targeted dredge sampling for *S.s. similis* in Long Island Sound and along the South shore of Long Island, focusing effort near inlets. Because no *S.s. similis* specimens were found south of Long Island, this project yielded only 13 *S.s. similis*. However, when combined with archived samples the total *S.s. similis* sample size was 134. Not all of these samples were included in the genomic analysis, either because of uneven collection sizes across localities or uneven sample quality (Table 1).

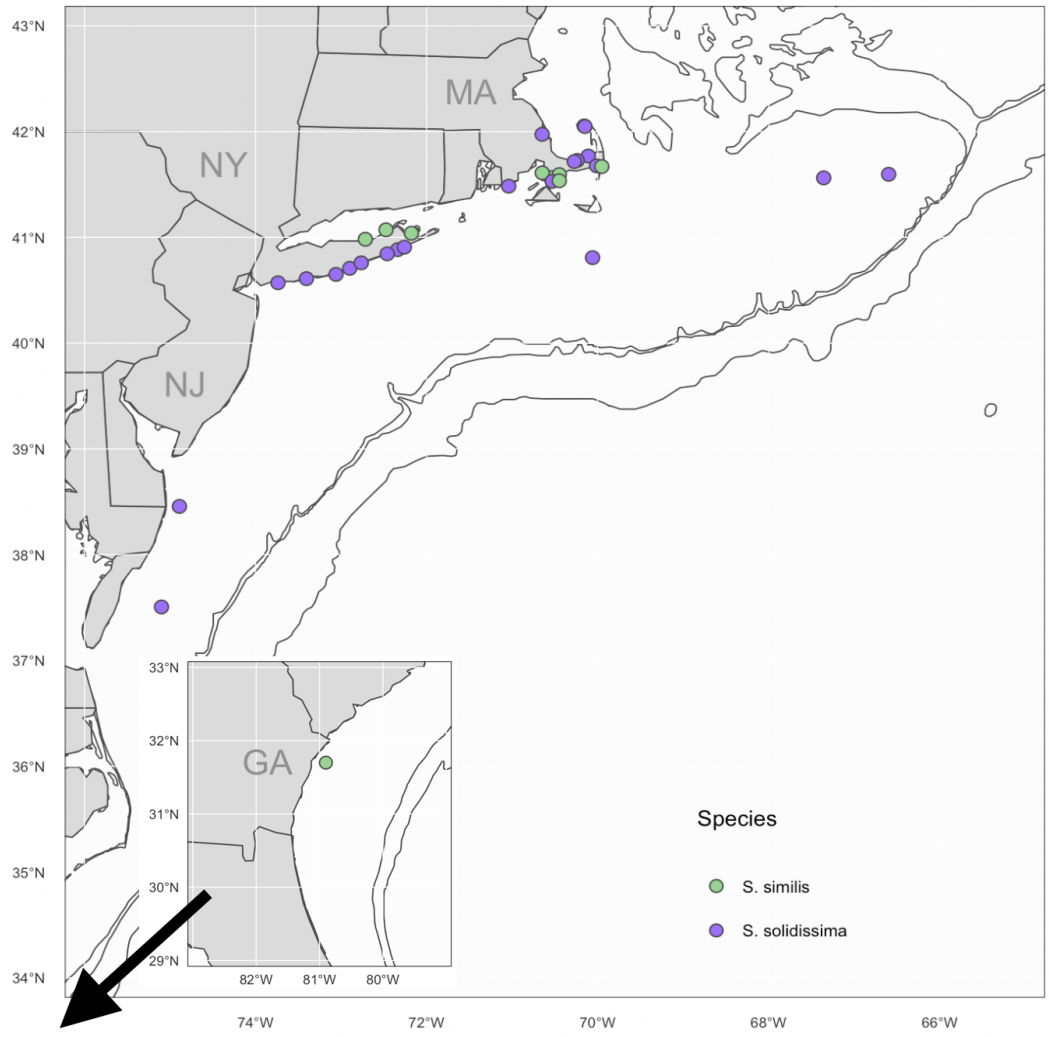

Figure 1: Sample collection locations for S.s. solidissima (purple) and S.s. similis (green) used in genomic analyses.

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 and therefore had fewer metadata. Additional 1999 federal samples archived by M. Hare also were analyzed. Closer to shore, samples included 2012 surfclams 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. Also, Georgia samples from 2012 were obtained using previous federal Hatch funds. The locations of all analyzed samples are shown in Fig. 1. Based on our sampling, mixed populations of these two nominal subspecies 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. The appendix Table A.2 lists other sites where dredges were attempted to find Long Island Spisula but none were discovered.

High Resolution Genomics Reveals an Additional Cryptic Taxon

Using a 'reduced representation' method of randomly sampling surfclam genomes, so that the same homologous chromosomal positions are sampled in each individual, we obtained a high resolution dataset consisting of 4.7 thousand quality-filtered single nucleotide polymorphisms (SNPs) from chromosomal loci scattered through the genome of both subspecies. This dataset consists of loci that have been carefully selected to be comparable (i.e. homologous) between *S.s. solidissima* and *S.s. similis* for all-inclusive analyses. Larger numbers of high quality loci and SNPs have been identified in each subspecies for separate, focused analyses (relevant data specified in figure captions; list of bioinformatic data sets in appendix).

In order to visualize multidimensional allele frequency variance among individuals in 2D space we used a principal component analysis (PCA). For this analysis we used only specimens sequenced in a single batch, including most *S.s. solidissima* samples and only representative samples from *S.s. similis* localities. SNPs were randomly subset to 2.6k SNPs for computational efficiency. With both subspecies included, the greatest variance in allele frequency among individuals is explained by the PC1 axis and separates *S.s. solidissima* from *S.s. similis* (Fig. 2). Surprisingly, *S.s. solidissima* samples also showed extensive allele frequency variance along PC2. This model-free statistical clustering of samples was the initial basis for identifying distinct *S.s. solidissima* operational taxonomic units (OTUs) A and B.

Phylogenetic inference was used to summarize evolutionary relationships among the main populations/OTUs. Using 2.1k SNP loci very stringently filtered for linkage disequilibrium (see methods: UPGMA) and Jukes-Cantor genetic distances that correct for multiple mutations at a given site, a UPGMA population tree clearly separates the nominal subspecies (Fig. 3). A phylogenetic tree for the same data, but analyzed at the individual instead of population level, can be seen in the appendix (Fig. A.3). In this phylogram the branch length between OTU A and the rest of *S.s. solidissima* is more than three times deeper (evolutionarily older) than all the diversity within *S.s. similis*. This relative age estimate may be an artifact of the analytical method - it is not reflected in a phylogeny constructed using a Bayesian MCMC approach that accounts for ancestral polymorphism. Different gene trees (inferred from markers at a particular chromosomal position) can disagree with each other and with the population tree due to the

diverse ways that ancestral polymorphisms sort out between diverging lineages. The model-based densitree (Fig. 4) uses MCMC to generate multiple trees drawn from the posterior distribution of species trees. It shows the same topology as the UPGMA tree, but no difference in the relative ages of the two subspecies. The grouping of populations/OTUs within the nominal subspecies clades will be discussed in their respective sections below.

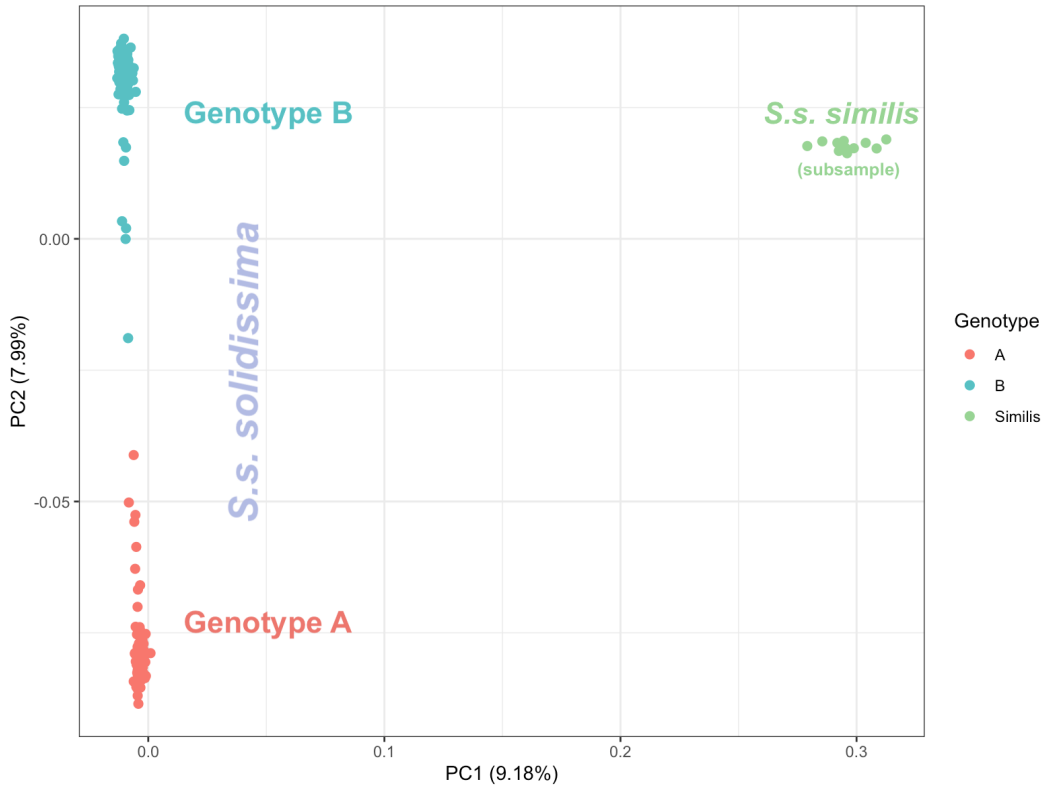

Figure 2: Principal components analysis plot of PC1 and PC2 summarizing allele frequency differentiation among individuals from both nominal subspecies. Genetically differentiated clusters of S.s. solidissima are labeled OTU Genotype A and OTU Genotype B. Created 2.6k SNPs found across both subspecies.

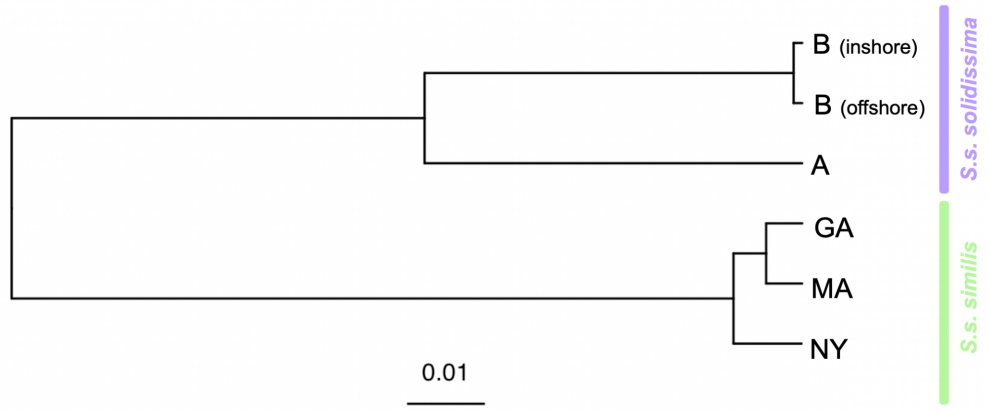

Figure 3: UPGMA (Unweighted Pair Group Method with Arithmetic mean) phylogenetic tree for Spisula solidissima sp. based on Nei's distance calculation of genetic distance between

populations based on 4.7k quality filtered and LD pruned SNPs. Branch lengths are scaled to represent the percentage of genetic variation. B offshore refers to samples collected from George's Bank, south of Nantucket, and New Jersey, while B inshore refers to all other OTU B samples.

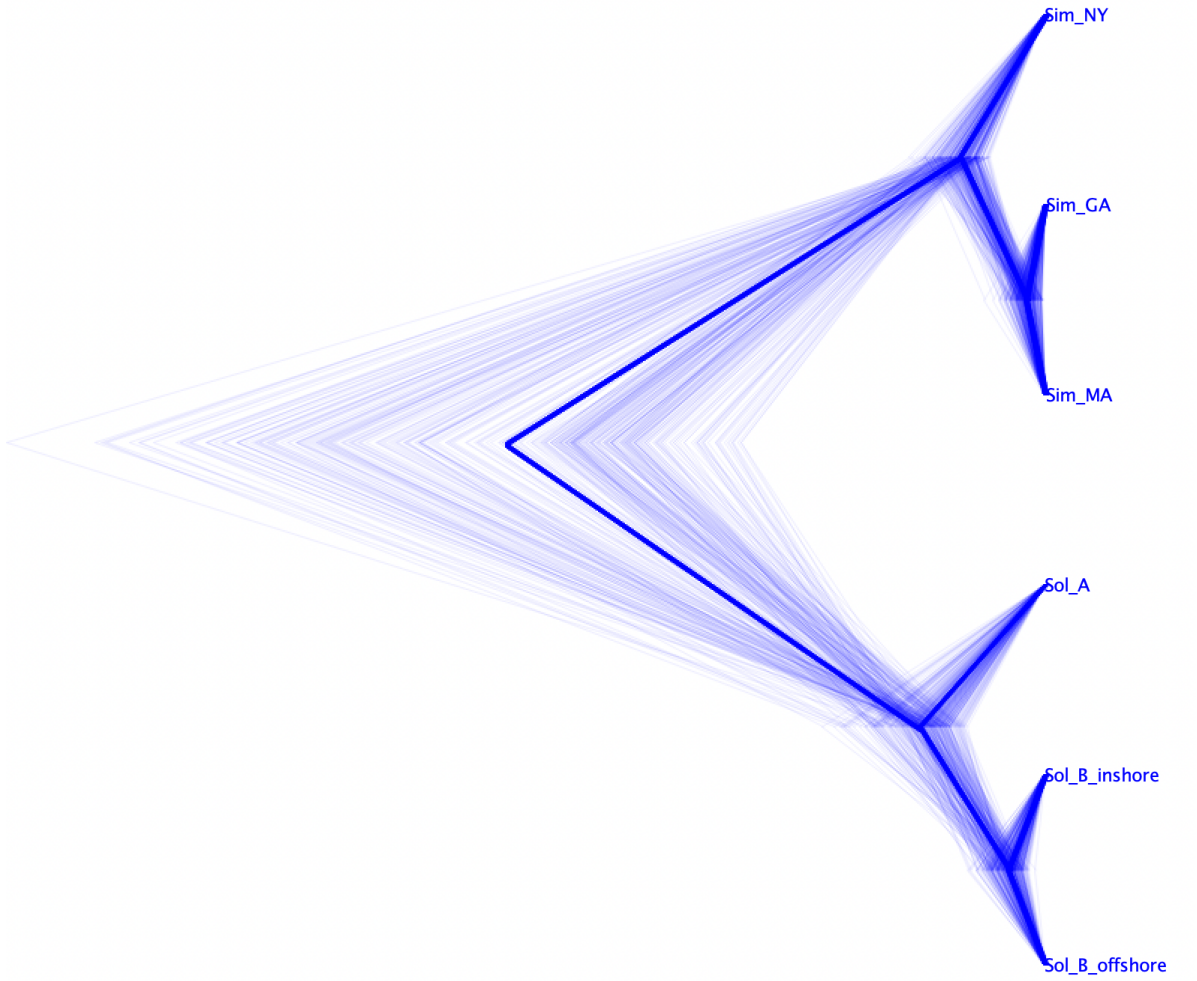

Figure 4: DensiTree depiction of SNAPP genetrees, where each faint blue line represents the most likely tree topology for a single marker. The thicker blue line illustrates the consensus tree. Inferred S.s. similis and S.s. solidissima clades as well as population topologies within subspecies are concordant with the neighbor joining tree (Fig. 3), but relative clade depth is only slightly greater in S.s. solidissima *in this result. S.s. soldissima individuals were split into OTUs as identified by PCA (Fig. 2) and geographic sampling region.*

S.s. similis **Population Structure and Connectivity**

Using PCA to explore patterns of population differentiation among 130 samples of *S.s. similis*, using 12.7 thousand filtered SNPs, three geographically discrete groups of samples show genetic differentiation (Fig. 5). The greatest differentiation along PC1 (2.88% allele frequency variance explained) separates Southern New England (NY+MA) from Georgia. Along PC2 the

differentiation is subtle between samples from the North shore of eastern Long Island and Peconic Bay (NY) versus the Southern coast of Massachusetts (MA), but it is interesting that there is any distinction at all between these geographically proximate populations. Using F_{ST} as a metric of allele frequency differentiation that spans from 0 to 1.0 (averaged among loci), the latitudinal contrast has F_{ST}=0.08 whereas between the two Southern New England populations F_{ST}=0.04.

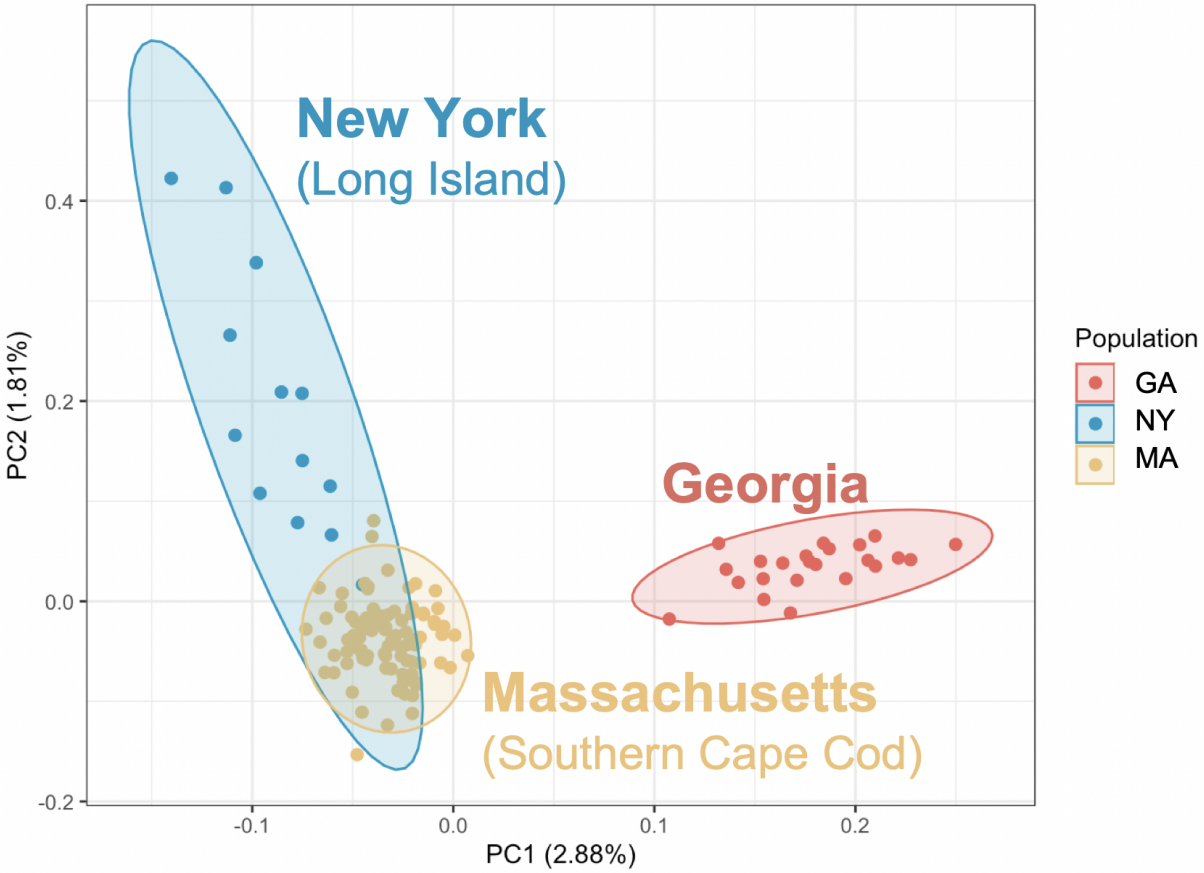

Figure 5: Principal component analysis of all S.s. similis *samples based on 12.7 thousand S.s. similis-specific SNPs. Each point represents a single individual where the distance between points illustrates genetic differences along the PC 1 and 2 axes. Percentages indicate the percentage of variation in the data set which each component axis accounts for.*

To assess population connectivity, we used two model-based estimators of gene flow; EEMS (Estimated Effective Migration Surfaces), and SpaceMix (Fig. 6). EEMS uses a pairwise genetic dissimilarity matrix at the individual level, derived from SNP data, to identify geographic regions where genetic similarity changes with distance. These levels of differentiation are then compared to an equilibrium expectation for isolation-by-distance (IBD), and regions of greater than expected genetic similarity or difference relative to IBD are depicted with colors. For example, two populations that share more genetic similarity than expected, given their geographic distance, will be shown in blue depicting greater-than-IBD levels of gene flow connectivity. SpaceMix clusters populations together in 2D space, similar to PCA but at the population rather than individual level. However, SpaceMix reports relative genetic differences

among populations in “geogenetic” space: a 2D positioning system based on both population allele frequencies and geographic sampling regions. The further a population in geogenetic space is from its geographic position, the more gene flow is inferred to deviate from equilibrium IBD expectations. In other words, a depiction of geogenetic population positions identical to geographic positions would indicate agreement with inferred IBD in which levels of gene flow are inversely proportional to geographic distance. In addition, SpaceMix also can estimate both the level of admixed alleles present in a population and where, in geogenetic space, those alleles are inferred to have originated from. Admixture from long distance migrants must have had a population source, but the source population was not necessarily sampled. Thus, geogenetic source inferences are based on the IBD model and accompanied by 95% confidence limits for inferred source position.

EEMS results for *S.s. similis* illustrate a small higher-than-IBD-predicted migration (blue) region north of Long Island, connecting the three sampling sites we had in that region. Otherwise, the most notable pattern is the low gene flow orange ‘barrier’ between New York (Long Island Sound) and both Massachusetts and Georgia populations. This pattern suggests that despite the geographic proximity of southern New England populations, MA and GA may exchange more migrants than the Long Island Sound population does with either of them. SpaceMix clusters all Massachusetts sites tightly with each other, and they partially overlap with New York samples, similar to the pattern observed with PCA. The admixture arrows depicted in Fig. 6b are quite faint (only the orange Georgia arrow is obvious), indicating a low percentage admixture inferred. The admixture origins for all northern sites are in and around other northern sites, whereas the origin of Georgia’s admixture is estimated to be from within the Massachusetts geogenetic ellipse (with a large geogenetic confidence limit that does not include the GA ellipse).

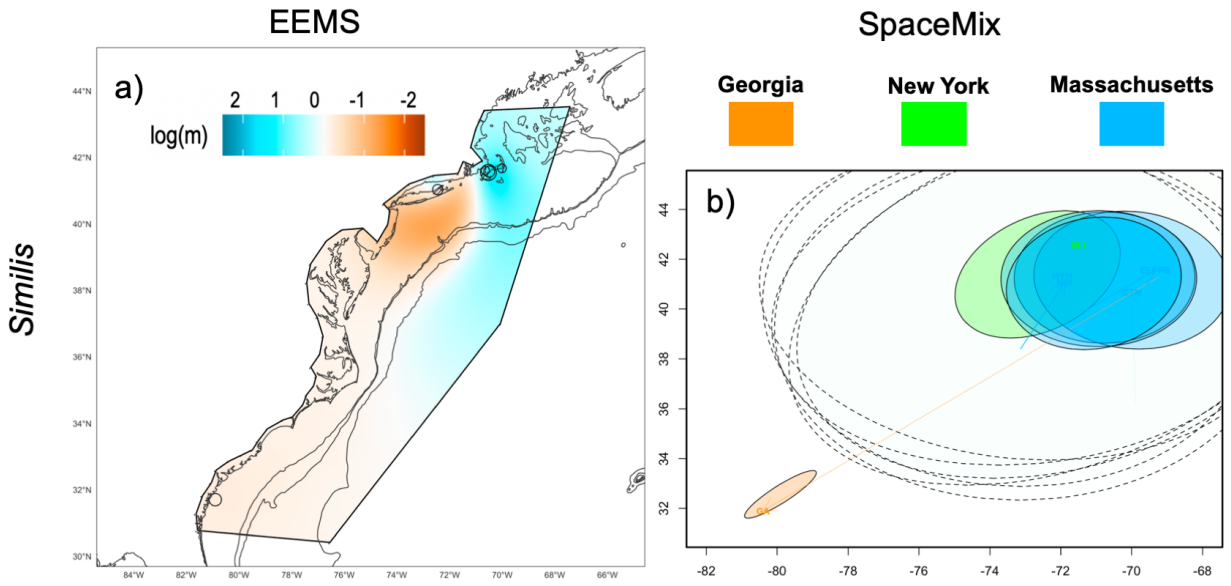

Figure 6: Estimates of population connectivity and gene flow. (a) EEMS estimates of connectivity are illustrated relative to the null hypothesis of isolation by distance shown in white (log(m) = 0). Darker orange areas show estimated lower-than-IBD gene flow whereas darker blue areas represent areas of relatively greater inferred gene flow. Black polygons describe the area in

which the EEMS analysis was conducted. Circles show the geographic centroid for sample locations combined into regional population samples for this analysis, scaled by the log of the number of samples in each regional sample. (b) Spacemix results for populations, colored by sampling region (regions defined as in EEMS analysis), are shown in "geogenetic space" which combines geographic sampling location inputs with genetic distance information. The axes show an adjusted "latitude" and "longitude" in geogenetic space rather than actual cartesian coordinates. The colored text labels show the geogenetic location estimate of each population, with the corresponding colored ellipses showing 95% confidence intervals on those estimates. Arrows depict the estimated source location and relative magnitude of long distance admixture in the population. Arrows originate at the estimated geogenetic source of long distance migrants, indicating long distance gene flow directionality toward the admixed population. Black dotted ellipses show 95% confidence intervals for the origin point of the admixture arrows. Larger versions of all panels can be found in the appendix.

The PCA and F_{ST} results make it surprising that Massachusetts and Georgia populations are sister taxa in both phylogenetic analyses. However, the EEMS analysis suggests that surfclams are relatively isolated within Long Island Sound. EEMS and SpaceMix results are consistent with gene flow between Massachusetts and Georgia, with directionality from North to South. Collectively, these analyses indicate Southern New England populations of *S.s. similis* have gene flow connectivity with populations to the south, but high East-West connectivity between Long Island Sound and Massachusetts should not be assumed. Zhang et. al (2015) predicted with a hydrodynamic biophysical model that *S.s. solidissima* larvae were more likely to recruit west and south of their spawning location along the Atlantic coast. Results here from *S.s. similis* (not modeled in the Zhang et al study), indicating the potential for long distance gene flow from Massachusetts south to Georgia, shows a possible agreement with the model predictions.

***S.s. solidissima* Population Structure and Connectivity**

Focusing analysis on *S.s. solidissima* yielded 49 thousand high quality SNPs. In a principle component analysis, OTUs A and B are largely partitioned along PC1 (13.3% of allele frequency variance explained), whereas PC2 shows some distinction within OTU (operational taxonomic unit) B between clams from Cape Cod Bay and the rest of the B OTU cluster (Fig. 7). Some Cape Cod Bay clams also show admixture between the A and B OTU clusters.

To examine admixture in more detail we used a model-based analysis that infers in each individual, for a given number of differentiated source populations contributing to the observed genotypic variation, what is the level of admixture from each source population (admixture is the genomic result of interbreeding with immigrants). The program STRUCTURE is typically run with various numbers of source populations assumed to determine with heuristics which provides the best explanation for the data. 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 more complicated than this and there are many

histories that could produce a 75/25 pattern in an individual. In general between two distinct source populations, individuals with moderate admixture (50/50) are likely to be initial hybrids, and minor admixture (90/10) may indicate an older history of hybrid backcrossing.

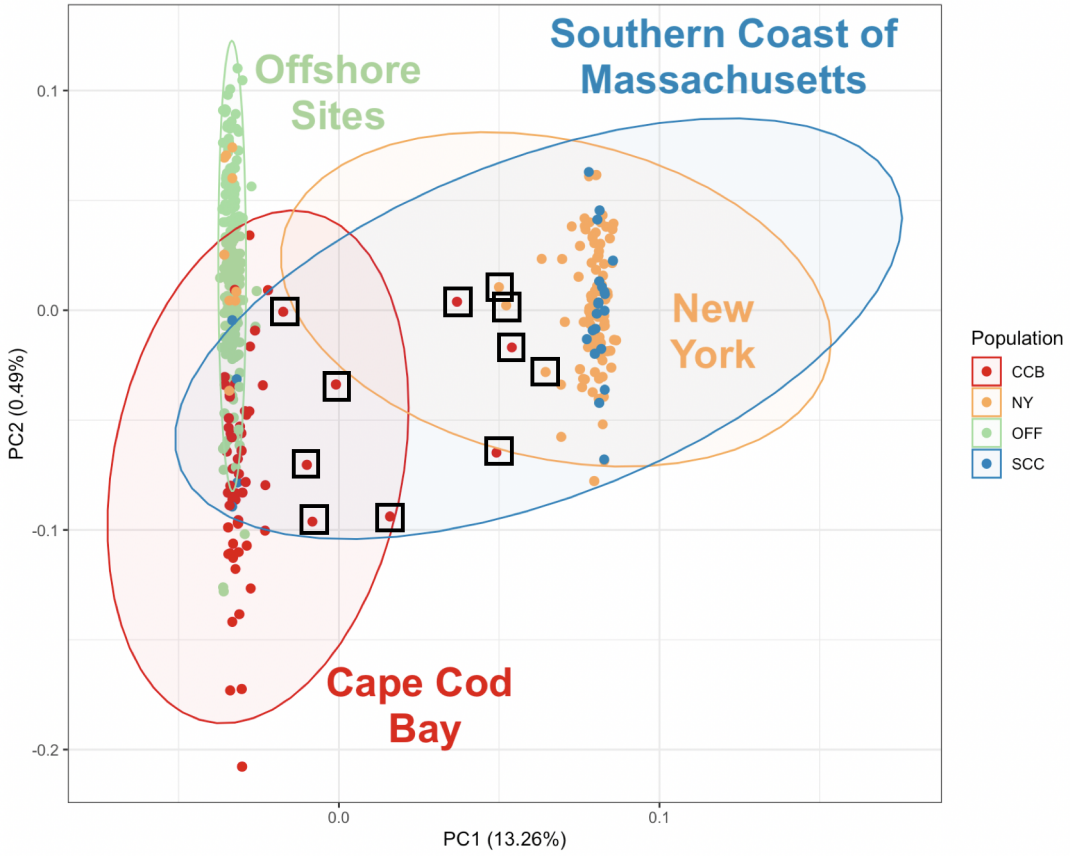

Figure 7: Principal component analysis of all S.s. solidissima *samples based on 49.7 thousand S.s. solidissima-specific SNPs. Offshore sites include samples from George's Bank, New Jersey and Nantucket. Black squares denote individuals shown in STRUCTURE to have admixture in 35-65% of genes (shown in Fig. 8-9).*

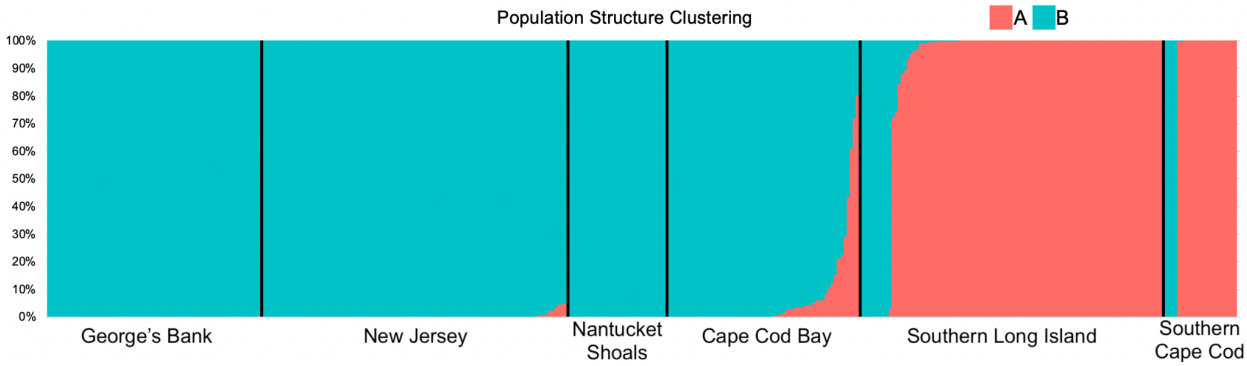

Figure 8: Clustering and admixture results from program STRUCTURE applied to all S.s. solidissima samples using 2.6 thousand haplotype loci. Models assuming K=2 source populations

showed the greatest support from the data, here depicted as blue-green for the B OTU source population and orange for the A OTU source population. Black vertical lines separate population samples and are just for reference. Each individual specimen is represented with a thin vertical bar that is either blue, orange, or a combination indicating proportional contributions from the two sources (admixture). In mixed populations, individual surfclams are arbitrarily ordered from fully type B to increasing proportions of type A.

As with the PCA, the greatest number of individuals admixed for OTUs A and B was found in Cape Cod Bay, but only a small minority of Cape Cod Bay surfclams (16%) had more than 10% admixture (Fig. 8-9). The admixed surfclams were scattered around Cape Cod Bay locations and no non-admixed type A surfclams were found, making the source of the OTU A admixture unclear. New Jersey was almost entirely B OTU, but also had slight admixture of A type in a few specimens. This is the only exception to offshore surfclams in federal waters being entirely B OTU. Southern Cape Cod and Southern Long Island were the only regions where nonadmixed A and B OTU surfclams co-occurred. Sympatry presents the opportunity for interbreeding, but A-B admixed clams only were found in Southern Long Island, not southern Cape Cod (Fig. 9).

We are making efforts to compare the depth distribution of A vs. B OTU surfclams, 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 depth and age data among samples makes interpretation difficult. Shells have been sent to the Woods Hole NOAA lab for aging to improve our ability to estimate von Bertalanffy growth curve parameters for both A and B OTUs (initial estimate of this growth curve shown in Appendix Fig. A.1).

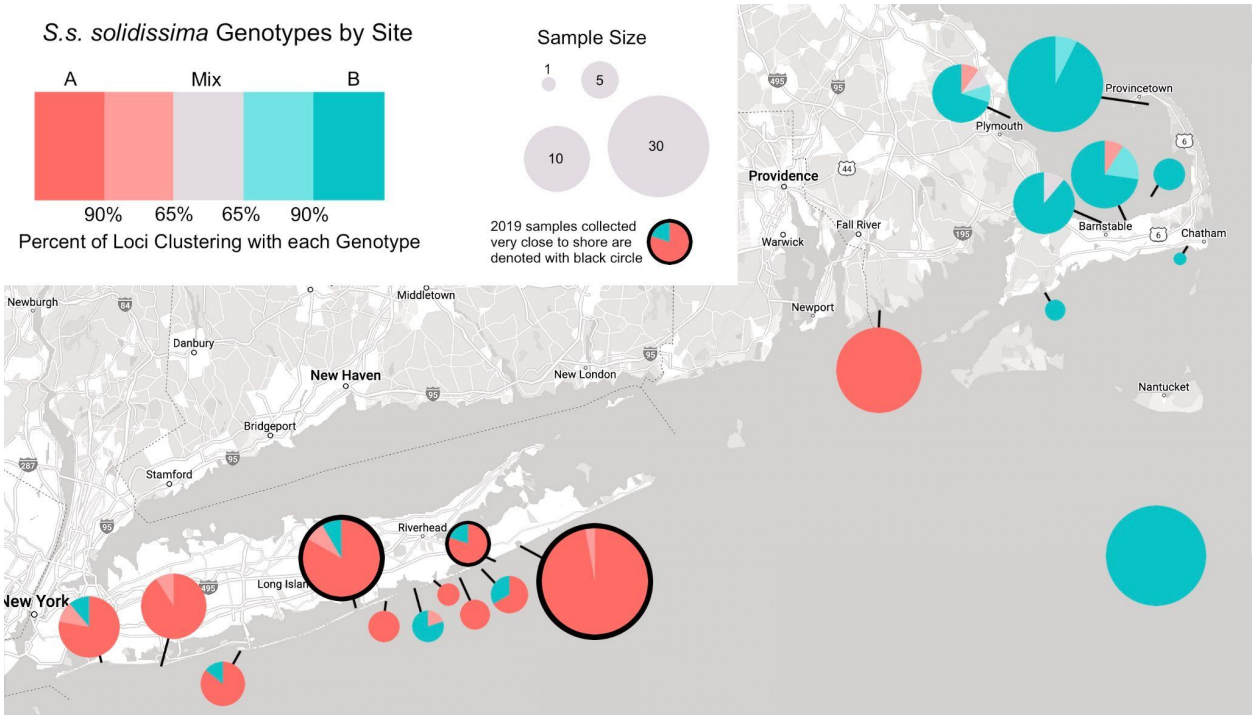

Figure 9: Southern New England pie diagrams depicting relative sample size and distribution of

OUTs A, B, and admixed individuals. Sites circled with black were collected in 2019 in very shallow water.

EEMS and SpaceMix were also used to assess population connectivity in *S.s. solidissima*. As the most evident population structure in *S.s. solidissima* is the separation between OTU A and B, we assessed the two OTUs separately and removed any individuals who showed moderate (10% or more) admixture between genotypes A and B.

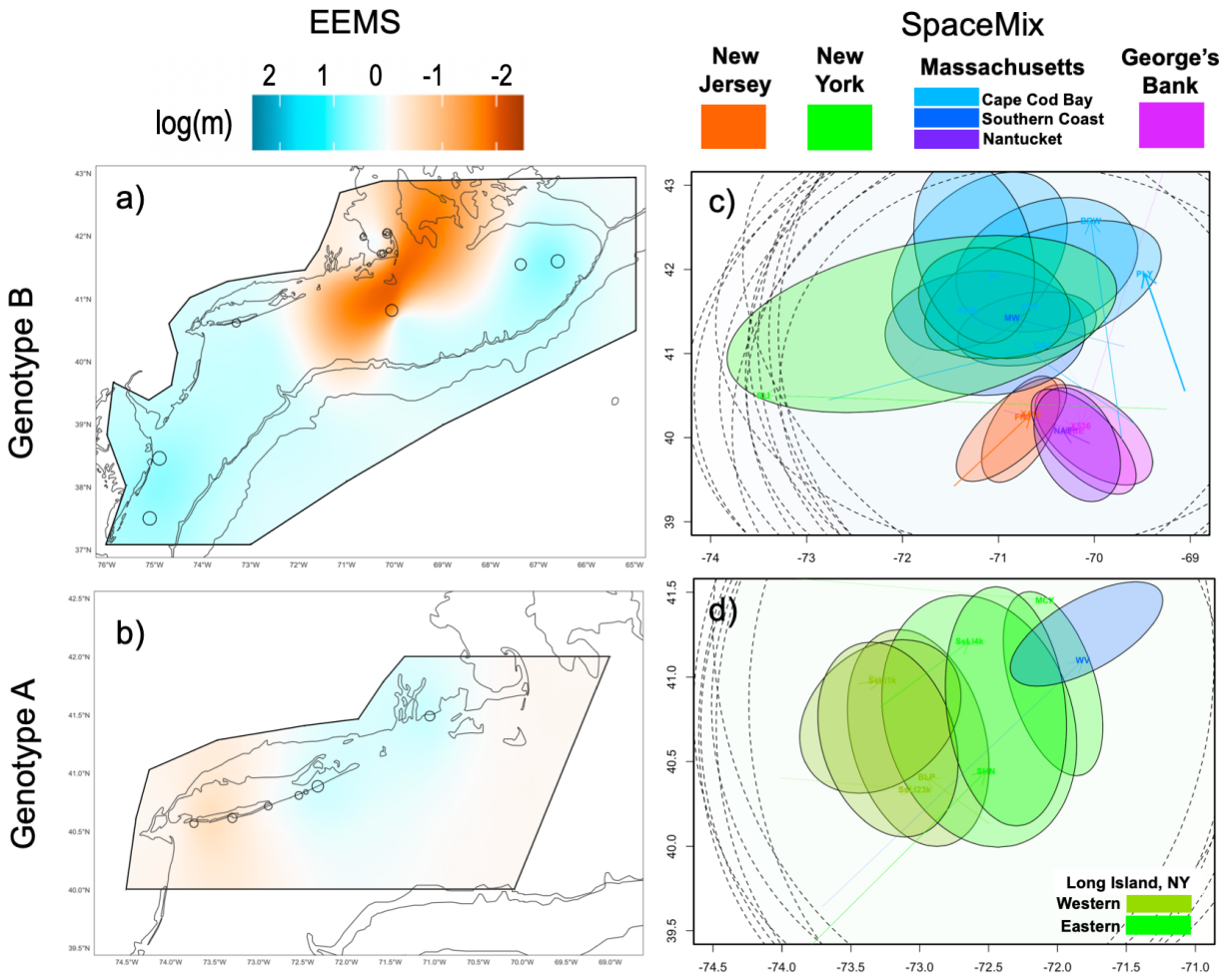

Figure 10: Estimates of population connectivity and gene flow. (a-b) EEMS estimates of connectivity are illustrated relative to the null hypothesis of isolation by distance shown in white (log(m) = 0). Darker orange areas show inferred regions of reduced migration whereas darker blue areas represent areas of greater migration. Black polygons describe the area in which the EEMS analysis was conducted. Circles show sampling locations (from Fig. 1) scaled by the log of the number of samples collected at each location. (c-d) Spacemix results for populations, colored by sampling region (Table A.1), are shown in "geogenetic space" which combines geographic sampling location inputs with genetic distance information. The axes show an adjusted "latitude" and "longitude" in geogenetic space rather than actual cartesian coordinates. The colored text labels show the geogenetic location estimate of each population, with the

corresponding colored ellipses showing 95% confidence intervals on those estimates. Arrows depict the inferred source region for long distance migration leading to admixture in a population, and the arrow thickness is proportional to the amount of admixture. Black dotted ellipses show 95% confidence intervals for the origin point of the admixture arrows. Larger versions of all panels can be found in the appendix.

For OTU B, IBD-level migration is inferred by EEMS within Cape Cod Bay (white color in Fig. 10a). However, there is an evident gene flow barrier between Cape Cod Bay and all offshore sites as well as southern New England (Fig. 10a). The inshore-offshore isolation also is represented in the SpaceMix results for OTU B (Fig. 10b). The inshore Massachusetts and New York population ellipses from SpaceMix cluster separately from those representing offshore sites from New Jersey and George's Bank. Nantucket Shoals, which is clustered with offshore sites in the PCA (Fig. 7) and SpaceMix, is represented as a region of reduced migration from both inshore and offshore locations in EEMS. A similar contradiction between EEMS and Spacemix was seen for the New York population of *S.s. similis* in Fig. 6. EEMS is perhaps over-extending the bounds of inferred gene flow boundary areas.

For OTU A, both EEMS and SpaceMix support slightly inflated gene flow between eastern Long Island and Massachusetts. The SpaceMix admixture arrows suggest that the majority of long distance migration for southern New England OTU A moves eastward, but it is not a prominent feature of the data as indicated by very large 95% confidence ellipses for the admixture source. In contrast, OTU A surfclams in western Long Island are inferred to experience lower levels of gene flow than expected under equilibrium IBD.

Genetic Diversity

To our surprise, all surfclam populations have similar levels of genetic diversity (Table 1; following page) 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 due to stronger genetic drift in a numerically small population. Allelic richness averaged near 3 for all three *S.s. similis* and *S.s. solidissima* regional populations while *S.s. solidissima* OTU A and OTU B populations averaged closer to 4.

Effective population size, N_e, is a key evolutionary parameter that is used in conservation biology for a variety of purposes ranging from assessing evolutionary potential to extinction risk. With our catalog of RADtags the best option for estimating N_e is the Moment-Based Temporal Method as calculated by NeEstimator 2.0 (Do, 2014). This requires genetic information from the same population across several generations and therefore we were only able to estimate effective population size at locations where we had archived samples to provide temporal comparison (shown in shades of gray in Table A.1). The *S.s. solidissima* populations we are able to estimate from are the offshore New Jersey and George's bank sites, for which we have samples from 1999 and 2020/2019. For *S.s. solidissima*, the generation time was estimated to

be six years (Weinberg 1999). Two and a half years was the estimate used for *S.s. similis* generation time based on an assumption of an age distribution similar to *S.s. solidissima*, scaled for *S.s. similis* max age of five year (Cerrato & Keith 1992; Jacobson & Old 1966). *S.s. similis* temporal N_e was only estimated for Massachusetts, where we calculated temporal N_e to be 314.9 (Jackknife confidence interval: 203.2-699.2). And for *S.s. solidissima* on George's Bank and in southern New Jersey, we calculated temporal N_e to be 923.6 (JK: 515.6-4427.6) and 664.0 (JK: 470.4-1128.4) respectively. Confidence intervals are notoriously large on estimates of contemporary N_e, but in this case the significantly smaller Ne for *S.s. similis* indicates that it experiences stronger genetic drift than *S.s. solidissima*, and this leads to more temporal change in allele frequency per generation. This finding should be replicated because gene flow and admixture can bias N_e estimates.

Table 1: Summary of sampled populations and diversity statistics. More detailed information on sampling can be found in the Appendix. OTU refers to "operational taxonomic unit" for the divisions we found in S.s. solidissima. N is the number of samples which were used in downstream analyses after filtering for samples with poor sequencing quality. Average values across all loci are reported ± standard deviation.

Population / OTU	**N**	**Haplotype Allelic Richness**	**Expected Haplotype Heterozygosity**
S.s. similis			
Georgia (GA)	23	3.11 ± 1.49	0.39 ± 0.24
New York (NY)	13	3.30 ± 1.57	0.42 ± 0.24
Massachusetts (MA)	88	3.25 ± 1.45	0.40 ± 0.22
***S.s. similis* Total**	124	6.34 ± 4.18	0.41 ± 0.22
S.s. solidissima			
George's Bank (GBE)	71	3.02 ± 1.36	0.34 ± 0.22
Cape Cod Bay (CCB)	60	3.07 ± 1.36	0.34 ± 0.22
Southern Cape (SCC)	24	3.24 ± 1.43	0.39 ± 0.22
Southern Long Island (SLI)	100	3.19 ± 1.37	0.37 ± 0.21
Nantucket (NAN)	30	3.04 ± 1.42	0.34 ± 0.22
New Jersey (NJ)	99	3.06 ± 1.36	0.34 ± 0.21
OTU A	113	3.93 ± 2.21	0.36 ± 0.22

OTU B Inshore	71	4.11 ± 2.21	0.34 ± 0.21
OTU B Offshore	200	4.06 ± 2.17	0.34 ± 0.22
***S.s. solidissima* Total**	384	7.35 ± 4.81	0.39 ± 0.20

Implications for Species and Connectivity

The patterns of genetic difference between *S.s. solidissima* and *S.s. similis* at hundreds of genes makes the case for full species-level classification. The phylogenetic trees presented here all corroborate the idea of reciprocal monophyly: that *S.s. solidissima* and *S.s. similis* both sort into their own clades which include no individuals from the other group. Additionally, even at sites where the nominal subspecies co-occur, we found no evidence of hybrids. Together these lines of evidence strongly support species level classification of *S.s. solidissima* and *S.s. similis* considering both the biological and phylogenetic species concept.

Within *S.s. similis*, the Massachusetts and New York populations show comparable levels of genetic diversity to the Georgia population, implying that they are not a recently migrated population suffering from founding effects. Additionally, it does appear that there is some level of gene flow from the northern populations, specifically Massachusetts, towards Georgia. Under the expectation that these northern populations were only small offshoots from a source population in the south, we would expect to see lower genetic diversity in the north and gene flow directed from the south towards Massachusetts and New York. Instead, these results suggest that gene flow may be directionally southward, supporting a biophysical and hydrodynamic model produced by Zhang et al. suggesting that *S.s. solidissima* larvae would move southward in this region (2015).

The most striking structure in *S.s. solidissima* is the genetic division found between OTUs A and B. Aside from a few hybrid individuals found in Cape Cod Bay, the *S.s. solidissima* OTUs are 2-3 times more genealogically distinct than the Georgia and northern populations of *S.s. similis*. However, genomic analyses across several modeling frameworks all consistently indicated moderate gene flow connectivity among populations within these OTUs. Strong connectivity between Georges Bank, the Delmarva shelf and Nantucket Shoals was observed for OTU B, and both OTUs presented moderate to high levels of connectivity at near shore sites. Genomic patterns and the occurrence of hybrids are consistent with a subspecies-level distinction between *S.s. solidissima* OTUs A and B. As OTU A individuals were found exclusively inshore, while OTU B occurs in both Massachusetts and on the shelf, previously observed differences in maximum size or other features of inshore *S.s. solidssima* could be related to these genetic differences. However, the size and age data available for assessment for this report (Fig. A.1) indicate similar growth and maximum size between OTUs A and B. Mixed populations of OTU A & B occurred along the South shore of Long Island and in southern Massachusetts, but hybrids were only in the former region.

In summary, this evidence supports consideration for the treatment of offshore OTU B (continental shelf populations), inshore OTU B (mostly Cape Cod Bay), and inshore OTU A populations as demographically separate populations, if not subspecies. Further investigation

should examine the sources of gene flow from OTU A into Cape Cod and investigate the depth distributions and life history differences between OTU A and B surfclams.

Methods

Sample Collection and DNA Sequencing

Samples used in this study were collected from gill tissue, adductor muscles, or small whole animals across a period of 21 years from 1999 to 2020 (Appendix Table A.1). Tissue samples were stored in ethanol until extraction.

DNA was extracted from gill or whole organisms using Qiagen DNeasy kit and adductor muscle samples were extracted with the E.Z.N.A.® Mollusc DNA Kit from Omega Bio-Tek. Extracted DNA was stored in a Tris elution at -20 °C. DNA quality was assessed using i) a Nanodrop ND-8000 spectrophotometer, and ii) 0.5% agarose gel electrophoresis to evaluate a subset of samples from each year and location for DNA degradation. DNA quantity was measured with a Qubit 2.0 Fluorometer.

Genome-wide sequencing was accomplished relatively cheaply for over 500 Spisula samples using double-digest Restriction site-Associated DNA sequencing (ddRAD-seq; Peterson et al., 2012) with *Pst*I and *Msp*I enzymes. ddRAD library preparation and DNA sequencing was performed by the University of Minnesota Genomics Center (UMGC) in three batches: i) pilot studies for each subspecies to determine a reasonable sequencing depth, ii) 135 *S.s. similis* individuals, and iii) 430 *S.s. solidissima* individuals with some overlapping individuals across batches to check for batch effects. Pooled ddRAD libraries were filtered to be between 450-600 base pairs (bp) long (=radtags, fragments of the genome defined by enzyme cut sites) before Illumina 150 bp paired-end sequencing with a NovaSeq platform.

Bioinformatics

Illumina reads were assessed for quality using fastqc and low quality sequences (with fewer than 1 million raw unique reads or with more than 50% missing data after an alignment) were removed from downstream analyses. Reads were then trimmed to remove adapters and assembled into a *de novo* reference catalog using dDocent (Puritz et al., 2014). A catalog is a collection of several thousand small contiguous sections of reads, called contigs (~300bp long), which a geneticist can use like a reference genome to map sequencing reads to when no reference exists for the species of interest, like in the case of Spisula. A separate catalog was created for each subspecies separately using 40 individuals from *S.s. similis* from across all localities, and 50 individuals from *S.s. solidissima* across all localities and operational taxonomic units (OTU). All analyses were conducted on data acquired by aligning sequence reads to the radtag catalog and calling single nucleotide polymorphisms (SNPs) within population samples. The SNPs were subjected to rigorous filtering as recommended by Puritz (O'Leary, 2018). From this pipeline, we modified the recommended max missing data per locus to be 0.9, allele balance threshold to be from 0.2 to 0.8, filtered at a max mean depth of 200, and filtered for sites deviating from HWE with a p value of 0.001.

To measure population structure, we analyzed either SNPs (two alleles per nucleotide site, filtered for LD) or sets of closely linked SNPs inferred to be a haplotype using rad_haplotyper (Willis et al., 2017). The benefit of haplotypes is that more than two alleles can

exist in a population, increasing the possibility of finding recently mutated alleles only in one population, thereby increasing spatial inferential power about gene flow based on this "private" genetic information. Allele frequency correlations among loci (linkage disequilibrium) can create artifactual population structure. Therefore, SNPs were filtered down to one SNP per radtag, keeping the SNP with the highest minor allele frequency. Linkage disequilibrium was removed from SNP and haplotype data by using plink 2.0 to prune loci with r^2>0.8, or r^2>0.2 (Chang et al., 2015). Haplotypes were used for STRUCTURE analyses and diversity statistics while SNP data were used for other analyses.

Table 2: Analyses used to assess population structure and diversity, and estimate gene flow and phylogenies. Analyses were conducted with filtered SNPs unless otherwise specified.

Analysis	**Purpose**	**Notes**
PCA	Illustrates multivariate allele frequency differentiation among individuals in 2D space. Provides model-free description of genetic clustering.	Calculated using the stats R package from LD (r^2<0.8) SNP data.
STRUCTURE	Model based estimate of genetic clusters with admixture proportions estimated per individual. For a given number of hypothesized distinct source populations contributing to genetic variation, cluster partition and admixture inferences are based on assumption of within-population Hardy-Weinberg equilibrium and linkage equilibrium. Allows us to estimate how many separate clusters our samples represent and how much genetic mixing there is between the clusters and which samples have the most mixing.	(Pritchard et al., 2000), separate runs assuming between 1-14 clusters (K), then selected the most likely number of clusters based on reported likelihood (in all cases K = 2 was the most likely). Calculated from haplotype data.
F_{ST} and AMOVA	Numerical values to describe the genetic difference between populations. Allow comparisons between across populations for relative genetic difference.	Calculated in Arlequin (Excoffier and Lischer, 2010) Calculated from LD (r^2<0.8) SNP data.
N_e (Effective Population Size)	N_e was estimated using temporal comparisons at sites where we had samples from multiple years (MA for *S.s. similis* and George's Bank and NJ for *S.s. solidissima*).	Estimated with NeEstimator v2 (Do et al., 2014) using a random subset of total LD-pruned SNPs (1260 markers for *S.s. similis*, 2540 makers for *S.s. solidissima*). We reported N_e

		based on only alleles with a minor allele frequency > 0.02 and used the Jorde/Ryman model for calculation. We used an estimate of 2 years for generation time for *S.s. similis* (approximating from Cerrato & Keith 1992; Jacobson & Old 1966) and 6 years for *S.s. solidissima* (Weinberg 1999).
Heterozygosity	Estimate of population-level diversity. The same number of individuals and loci were used across subspecies for comparability: 39 samples (13 per population), and 1050 haplotype loci.	Calculated in Arlequin using haplotype data. Total heterozygosity was calculated across all samples as one, not an average across populations.
Haplotype Allele Richness	Alternate estimate of population-level diversity. The same number of individuals and loci were used across subspecies for comparability: 39 samples (13 per population), and 1050 haplotype loci.	Calculated with R packages adegenet and hierfstat (Goudet, 2005; Jombart, 2008) using haplotype data. Haplotype loci were subset in solidissima to have equal numbers of loci and individuals to make data sets comparable for comparison.
EEMS	Estimated Effective Migration Surfaces identifies geographic regions where genetic similarity changes with distance and models a 2D mesh of population structure and migration estimates using this geographic information relative to the IBD model.	(Petkova et al., 2016), EEMS parameters were set to the following for all analyses: --nDemes 700 --numMCMCIter 2000000 --numBurnIter 1000000 --numThinIter 999
SpaceMix	Uses geographic and genetic information to create a map in “geogenetic” space, informative about deviations from an equilibrium isolation by distance expectations. The amount and source of long distance migration and admixture can be estimated for each population.	(Bradburd et al. 2016) SpaceMix parameters were set to the following for all analyses: n.fast.reps = 10, fast.MCMC.ngen = 1e5, ngen = 1e6, mixing.diagn.freq = 50 All analyses presented here were performed with the source and target model which had comparable likelihood to all other methods across all data sets.

SNAPP	Applies a multispecies coalescent model to infer phylogenetic relationships among OTUs for many ddRAD loci assumed to be independent (freely recombining, i.e. low linkage disequilibrium). OTU membership was pre-determined, and was based on PCA and STRUCTURE results.	(Bryant et al., 2012), performed with LD (r^2>0.8) SNP dataset. SNAPP was run over 8 million MCMC chains using a gamma distribution to estimate lambda with (alpha = 2.0, beta = 200.0) and otherwise default SNAPP parameters as set up with Beauti in Beast version 2.6.7. Vcf files were converted to nexus files using vcf2phylip.py prior to use in Beauti
UPGMA Tree: R packages adgenet, phangorn, and ggtree	Create UPGMA (Unweighted Pair Group Method with Arithmetic mean) phylogenetic trees at the population level from SNPs. This helps illustrate the difference between populations at a species/evolutionary scale.	(Jombart, 2008; Schliep et al., 2016, Yu et al., 2017), performed with LD (r^2<0.2) SNP data. For the individual-level tree, the distance matrix is calculated with Nei's distance metric using vcf2PopTree. For population-level tree, distance was calculated with dist.genpop from the adgenet package.
Vcf2PopTree	Create neighbor-joining phylogenetic trees at the individual level from SNPs. Branch lengths are proportional to genetic differences.	(Subramanian, 2019), performed with LD (r^2<0.2) SNP dataset

Key Project Objectives and Deliverables

1. 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.
    a. High quality transcriptomes were assembled for each subspecies and will be reported in a peer-reviewed journal, with sequence data and assemblies deposited into public databases. These will represent the first publicly available *Spisula* transcriptomes, and nearly the first from the Mactridae bivalve family.
2. 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.
    a. Considerable effort was put into this objective based on DNA polymorphisms in the transcriptomes (while the pandemic delayed completion of the genomic sequencing from larger sample sizes). There was some success, but when a dozen markers were tested on larger sample sets, none proved to be diagnostic. Two dozen apparently diagnostic markers have been bioinformatically identified based on the genomic sequence data, but laboratory testing of these has not yet been possible.
3. 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.
    a. This contract with Matt Weeks was very successful, both for obtaining samples from Long Island, NY, but also for definitively ruling out the South coast of Long Island as *S.s. similis* habitat.
4. 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.
    a. Because genomic sequence data were so delayed by the pandemic, full testing of diagnostic markers and application to all existing samples was not possible. In addition, both state and federal sampling was curtailed by the pandemic, leaving us with many fewer than 3000 samples total (Appendix 1).
5. Collect genome-scale data from 350 samples and identify DNA variants within and between each subspecies.
    a. Fully accomplished and described in this report.
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.
    a. Fully accomplished and described in this report.

Literature Cited

Ambrose Jr, W. G., D. S. Jones, and I. Thompson. "Distance from shore and growth rate of the suspension feeding bivalve, Spisula solidissima." *Proceedings of the National Shellfisheries Association*. Vol. 70. No. 2. 1980.

Bradburd, Gideon S., Peter L. Ralph, and Graham M. Coop. "A spatial framework for understanding population structure and admixture." *PLoS genetics* 12.1 (2016): e1005703.

Cerrato, Robert M., and Debra L. Keith. "Age structure, growth, and morphometric variations in the Atlantic surf clam, Spisula solidissima, from estuarine and inshore waters." *Marine Biology* 114.4 (1992): 581-593.

Bryant, David, et al. "Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis." *Molecular biology and evolution* 29.8 (2012): 1917-1932.

Chang, Christopher C., et al. "Second-generation PLINK: rising to the challenge of larger and richer datasets." *Gigascience* 4.1 (2015): s13742-015.

Do, Chi, et al. "NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data." *Molecular ecology resources* 14.1 (2014): 209-214.

Excoffier, Laurent, and Heidi EL Lischer. "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." *Molecular ecology resources* 10.3 (2010): 564-567.

Hare, Matthew P., and Harmony Borchardt-Wier. "Microsatellite primers designed from pyrosequences to compare two subspecies of surf clam, Spisula solidissima." *Conservation genetics resources* 6.4 (2014): 1039-1041.

Hare, Matthew P., and James R. Weinberg. "Phylogeography of surfclams, Spisula solidissima, in the western North Atlantic based on mitochondrial and nuclear DNA sequences." *Marine Biology* 146.4 (2005): 707-716.

Hare, Matthew P., et al. "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 (2010): 799-807.

Hardy, Olivier J., and Xavier Vekemans. "Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models." *Heredity* 83.2 (1999): 145-154.

Goudet, Jérôme. "Hierfstat, a package for R to compute and test hierarchical F-statistics." *Molecular ecology notes* 5.1 (2005): 184-186.

Jacobson, M. K., and W. E. Old. "On the identity of *Spisula similis* (Say)." *Am. Malacol. Union Rep.* (1966):30–31.

Jombart, Thibaut. "adegenet: a R package for the multivariate analysis of genetic markers." *Bioinformatics* 24.11 (2008): 1403-1405.

Jones, Douglas S., Ida Thompson, and William Ambrose. "Age and growth rate determinations for the Atlantic surf clam Spisula solidissima (Bivalvia: Mactracea), based on internal growth lines in shell cross-sections." *Marine Biology* 47.1 (1978): 63-70.

Jones, Douglas S. "Annual cycle of shell growth increment formation in two continental shelf bivalves and its paleoecologic significance." *Paleobiology* 6.3 (1980): 331-340.

O'Leary, Shannon J., et al. "These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists." (2018): 3193-3206.

Peterson, Brant K., et al. "Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species." *PloS one* 7.5 (2012): e37135.

Petkova, Desislava, John Novembre, and Matthew Stephens. "Visualizing spatial population structure with estimated effective migration surfaces." *Nature genetics* 48.1 (2016): 94-100.

Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. "Inference of population structure using multilocus genotype data." *Genetics* 155.2 (2000): 945-959.

Puritz, Jonathan B., Christopher M. Hollenbeck, and John R. Gold. "dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms." *PeerJ* 2 (2014): e431.

Ropes, J. W. "Biological and fisheries data on surf clam, Spisula solidissima (Dillwyn, 1817)." *US DOC, tech. ser. rep* 24 (1980).

Schliep, Klaus, et al. *Intertwining phylogenetic trees and networks*. No. e2054v1. PeerJ Preprints, 2016.

Subramanian, Sankar, Umayal Ramasamy, and David Chen. "VCF2PopTree: a client-side software to construct population phylogeny from genome-wide SNPs." *PeerJ* 7 (2019): e8213.

Walker, Randal L. "Comparative gametogenesis of Spisula solidissima solidissima and Spisula solidissima similis cultured in coastal Georgia." *Journal of the World Aquaculture Society* 29.3 (1998): 304-312.

Weinberg, J. R. "Age-structure, recruitment, and adult mortality in populations of the Atlantic surfclam, Spisula solidissima, from 1978 to 1997." *Marine biology* 134.1 (1999): 113-125.

Willis, Stuart C., et al. "Haplotyping RAD loci: an efficient method to filter paralogs and account for physical linkage." *Molecular Ecology Resources* 17.5 (2017): 955-965.

Yu, Guangchuang, et al. "ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data." *Methods in Ecology and Evolution* 8.1 (2017): 28-36.

Zhang, Xinzhong, et al. "Atlantic surfclam connectivity within the Middle Atlantic Bight: mechanisms underlying variation in larval transport and settlement." *Estuarine, Coastal and Shelf Science* 173 (2016): 65-78.

Zhang, Xinzhong, et al. "Modeling larval connectivity of the Atlantic surfclams within the Middle Atlantic Bight: model development, larval dispersal and metapopulation connectivity." *Estuarine, Coastal and Shelf Science* 153 (2015): 38-53.

Appendix

Appendix Tables

Appendix Figures

Table A.1: Full list of sampling locations and the number of individuals sequenced from each location. Pairs of sampling sites used for temporal comparisons to calculate N_e are highlighted in gray. Temporal N_e was only estimated in Southern Cape Cod for S.s. similis.

Location	Year	Lat	Lon	Average Depth (ft)	Region ID (description)	Number of Samples Sequenced	
						S.s. similis	*S.s. solidissima*
Nantucket Shoals, MA	2020	40.82	-70.06	-	NAT (Nantucket)	-	30
Fenwick Island, NJ	2020	38.46	-74.90	-	NJ (New Jersey)	-	49
New Jersey	1999	37.50	-75.10	29.0	NJ	-	26
George's Bank East	2019	41.60	-66.60	-	GBE (George's Bank East)	-	45
George's Bank East	1999	41.55	-67.37	53.5	GBE	-	50
Mecox Bay, NY	2019	40.89	-72.33	17.4	SLI (Southern Long Island)	-	34
Shinnicock, NY	2019	40.84	-72.46	18.8	SLI	-	5
Bellport, NY	2019	40.72	-72.89	11.4	SLI	-	12
Goldsmith Inlet, NY	2019	41.07	-72.47	12.8	NLI (Northern Long Island)	8	-
Smithtown, NY	2019	40.99	-72.70	18.8	NLI	2	-
Peconic Bay, NY	2019	41.04	-72.19	10	NLI	3	-
Buzzards Bay, MA	2017	41.49	-71.05	21.5	CCB (Cape Cod Bay)	-	20
Cape Cod Bay, MA	2017	41.77	-70.11	5.2	CCB	-	5

Nantucket Sound, MA	2017	41.67	-70.02	2.5	SCC (Southern Cape Cod)	9	1
Provincetown Harbor, MA	2017	42.06	-70.16	1.5	CCB	-	11
Chase Garden Creek, MA	2017	41.72	-70.24	2.5	CCB	-	11
Plymouth Bay, MA	2017	41.98	-70.65	2.5	CCB	-	10
Barnstable Harbor, MA	2017	41.72	-70.28	3.0	CCB	-	10
Eel Pond, MA	2017	41.55	-70.55	3.5	SCC	10	-
West Falmouth Harbor, MA	2017	41.61	-70.65	3.5	SCC	8	-
Popponesset Bay, MA	2017	41.59	-70.45	3.5	SCC	6	-
Washburn Island, MA	2012	41.55	-70.53	-	SCC	56	3
Provincetown, MA	2012	42.04	-70.15	-	CCB	-	16
Long Island Shelf, NY	2012	40.77	-72.66	33.3	SLI	-	49
Georgia	2012	31.70	-80.90	-	GA (Georgia)	23	-

Total Sequenced Samples	***S.s. similis***	***S.s. solidissima***
	125	**391**

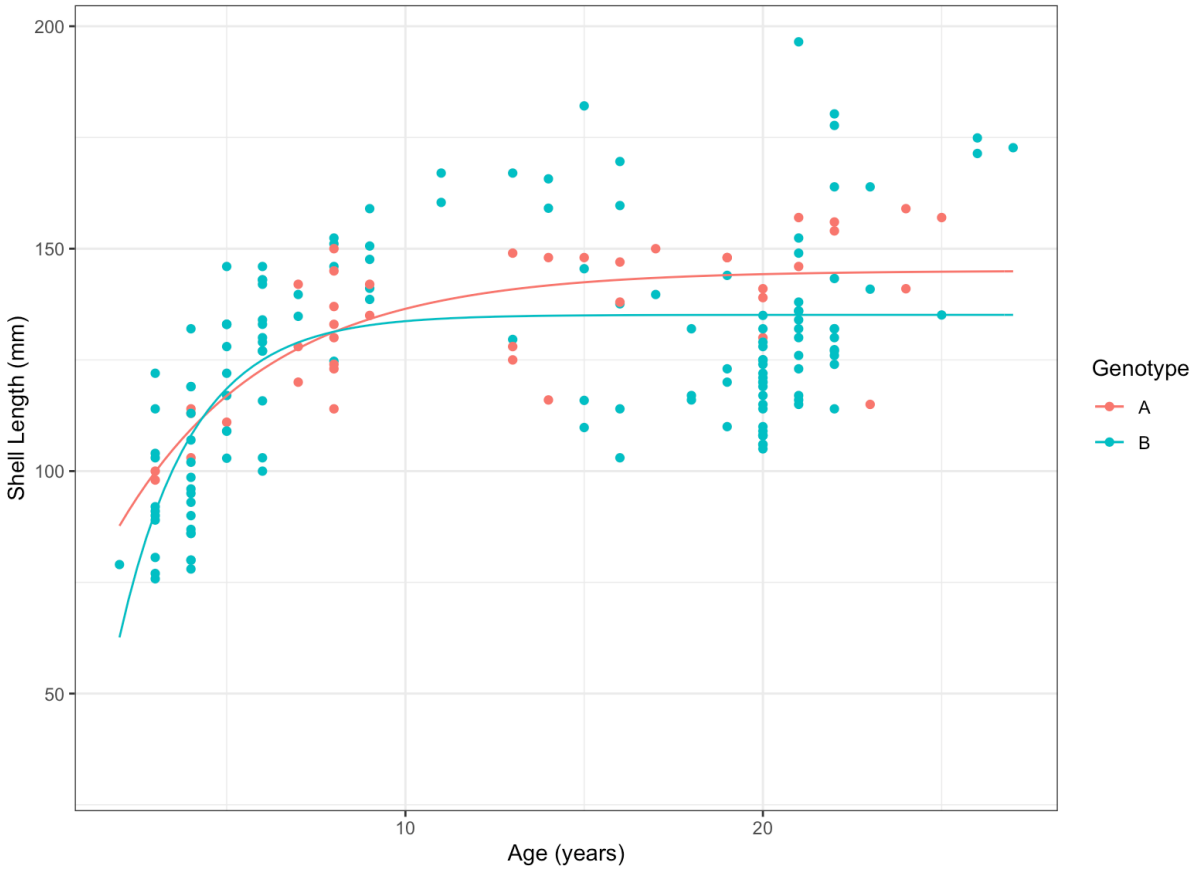

Figure A.1: The von Bertalanffy growth curve estimates for OTU A (orange) and B (blue).

This is an estimation of Von Bertalanffy growth curves for OTU A and B. These graphs should be considered preliminary and conclusions should be drawn after more data can be added after we receive age data from Wood's Hole on further samples.

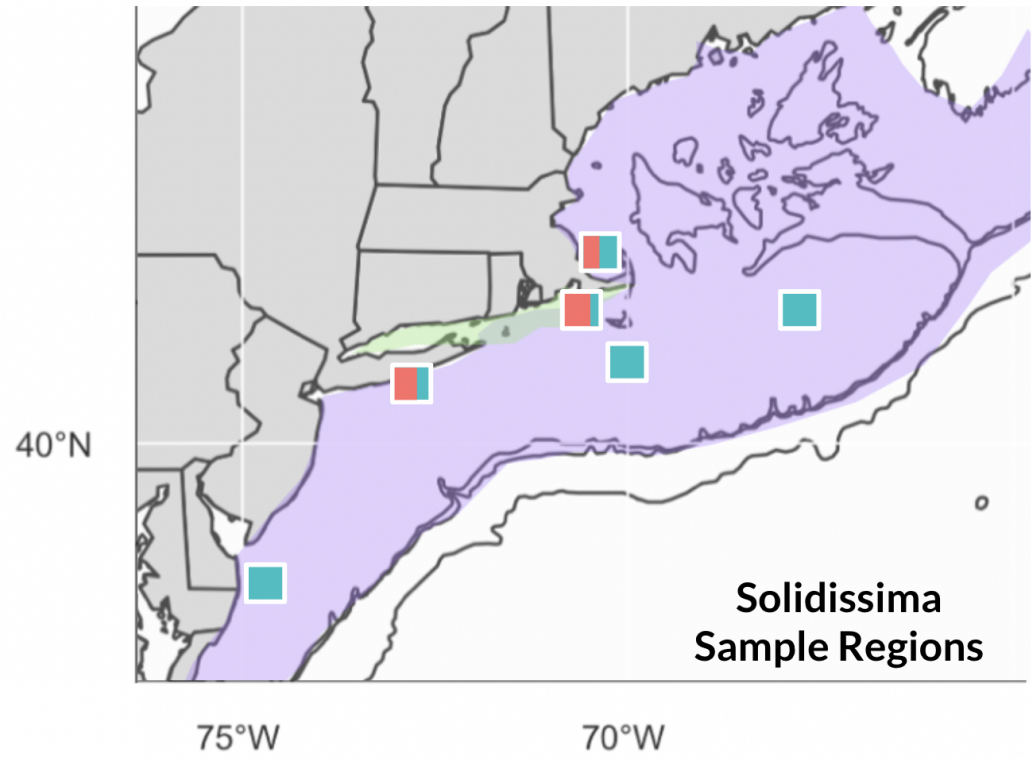

Figure A.2: The proportion of OTU A (orange) and OTU B (blue) in the sampled regions. The purple area represented reported range of S.s. solidissima, while green notates the northern range of S.s. similis (Hare and Weinburg, 2005).

This figure is intended to help provide context for Fig A.1, illustrating the regions where the individuals of each OTU were found.

Table A.2: Dredge Locations sampled by Matt Weeks along Long Island in August 2019. Each tow is represented with two rows in the table, indicating the information at the start and end of the tow. The catch is the total number of individuals gathered from that tow and appears the same in the Start and End rows for each tow. Additional information on which samples were collected from each site and their OTU designation if sequenced can be found in Table A.4.

Tow	Start/ End	Station	Station Name	Lat	Long	Date	Time	Temp	Depth	Speed (kt)	Catch	Notes
1	Start	1	Goldsmith Inlet	41.06976047	-72.47173316	2019-08-12	2:12:28 PM	86	11	2	0	no liner, low tide
1	End	1	Goldsmith Inlet	41.07020312	-72.4699313	2019-08-12	2:21:46 PM	86	11	3	0	no liner
2	Start	1	Goldsmith Inlet	41.0678323	-72.47297821	2019-08-12	2:53:10 PM	86	11	3	1	no liner
2	End	1	Goldsmith Inlet	41.06932428	-72.46981521	2019-08-12	2:58:01 PM	86	11	1.5	1	no liner
5	End	1	Goldsmith Inlet	41.0705275	-72.47778129	2019-08-12	4:02:59 PM	86	16	2	0	no liner, slack low
5	Start	1	Goldsmith Inlet	41.07288089	-72.47260194	2019-08-12	4:10:08 PM	86	9	3	0	no liner
6	End	1	Goldsmith Inlet	41.06661743	-72.48360235	2019-08-12	4:29:29 PM	86	16	1.5	0	no liner
6	Start	1	Goldsmith Inlet	41.06757548	-72.4821144	2019-08-12	4:34:21 PM	86	15	1.5	0	no liner
7	End	1	Goldsmith Inlet	41.07049548	-72.47725457	2019-08-12	4:40:33 PM	86	14	1.5	1	no liner
7	Start	1	Goldsmith Inlet	41.06843311	-72.48066726	2019-08-12	4:56:53 PM	86	19	1.5	1	no liner
8	Start	1	Goldsmith Inlet	41.07176961	-72.47621748	2019-08-12	5:05:48 PM	86	17	1.5	0	no liner
8	End	1	Goldsmith Inlet	41.06093458	-72.49541162	2019-08-12	5:39:30 PM	87	19	1.5	0	no liner
9	Start	1	Goldsmith Inlet	41.07940754	-72.45572565	2019-08-13	8:23:31 AM	86	16	1.5	0	no liner, high tide, sand hills
9	End	1	Goldsmith Inlet	41.07335279	-72.46401669	2019-08-13	8:40:45 AM	85	13	1.5	0	no liner, rocky
10	Start	1	Goldsmith Inlet	41.0798237	-72.45589438	2019-08-13	9:27:30 AM	85	17	1.5	0	no liner
10	End	1	Goldsmith Inlet	41.0740888	-72.46421291	2019-08-13	9:48:28 AM	85	14	1.5	0	no liner
11	Start	1	Goldsmith Inlet	41.07426365	-72.45602321	2019-08-13	10:14:38 AM	85	13	1.5	0	no liner
11	End	1	Goldsmith Inlet	41.06726887	-72.46214946	2019-08-13	10:50:39 AM	85	0	1.5	0	no liner
12	Start	1	Goldsmith Inlet	41.06236604	-72.47381849	2019-08-13	11:15:15 AM	85	18	1.5	0	no liner
12	End	1	Goldsmith Inlet	41.05717613	-72.48279375	2019-08-13	11:38:44 AM	85	15	1.5	0	no liner

13	Start	1	Goldsmith Inlet	41.05437406	-72.48430065	2019-08-13	12:46:21 PM	86	13	1.5	2	liner, razor clams
13	End	1	Goldsmith Inlet	41.05357585	-72.4885626	2019-08-13	1:04:29 PM	86	0	1.5	2	liner, deckers
14	Start	1	Goldsmith Inlet	41.05614977	-72.48599698	2019-08-13	1:27:02 PM	86	17	1.5	0	liner
14	End	1	Goldsmith Inlet	41.05046726	-72.49415156	2019-08-13	1:53:53 PM	86	16	1.5	0	liner
15	Start	1	Goldsmith Inlet	41.05525458	-72.48485109	2019-08-13	2:07:33 PM	86	12	1.5	0	liner
15	End	1	Goldsmith Inlet	41.05026115	-72.49366357	2019-08-13	2:37:41 PM	86	0	1.5	0	liner
16	Start	1	Goldsmith Inlet	41.05512114	-72.50001152	2019-08-13	2:51:04 PM	86	0	1.5	0	liner
16	End	1	Goldsmith Inlet	41.06009832	-72.49246881	2019-08-13	3:14:35 PM	86	0	1.5	0	liner
17	Start	1	Goldsmith Inlet	41.06352291	-72.4912132	2019-08-13	3:26:53 PM	86	19	1.5	2	liner
17	End	1	Goldsmith Inlet	41.05994686	-72.48679904	2019-08-13	3:38:57 PM	86	10	1.5	2	liner
18	Start	1	Goldsmith Inlet	41.06526366	-72.48745903	2019-08-13	3:51:28 PM	86	21	1.5	0	liner
18	End	1	Goldsmith Inlet	41.06170303	-72.48404676	2019-08-13	4:11:41 PM	86	0	1.5	0	liner
19	Start	1	Goldsmith Inlet	41.06158837	-72.49620094	2019-08-13	4:23:10 PM	86	19	1.5	0	liner
19	End	1	Goldsmith Inlet	41.05397282	-72.49107616	2019-08-13	4:45:41 PM	86	14	1.5	0	liner
20	Start	1	Goldsmith Inlet	41.05354467	-72.49568327	2019-08-13	4:56:31 PM	86	10	1.5	1	liner
20	End	1	Goldsmith Inlet	41.05808892	-72.48604861	2019-08-13	5:18:43 PM	86	11	1.5	1	liner
21	Start	1	Goldsmith Inlet	41.05733807	-72.48579372	2019-08-13	5:25:54 PM	86	14	1.5	0	liner
21	End	1	Goldsmith Inlet	41.05958828	-72.48094738	2019-08-13	5:43:33 PM	86	0	1.5	0	liner
22	Start	1	Goldsmith Inlet	41.05981811	-72.48140864	2019-08-13	5:46:32 PM	86	11	1.5	1	liner
22	End	1	Goldsmith Inlet	41.06483786	-72.47253011	2019-08-13	6:11:38 PM	86	13	1.5	1	liner
23	Start	2	Peconic Bay	41.03559298	-72.19355799	2019-08-14	7:08:24 AM	86	16	1.5	0	liner
23	End	2	Peconic Bay	41.0373127	-72.19157525	2019-08-14	7:20:05 AM	86	0	1.5	0	liner
24	Start	2	Peconic Bay	41.03848348	-72.19073539	2019-08-14	7:30:32 AM	86	14	1.5	0	liner
24	End	2	Peconic Bay	41.03882337	-72.1852698	2019-08-14	7:40:46 AM	85	14	1.5	0	liner
25	Start	2	Peconic Bay	41.03817461	-72.18813499	2019-08-14	7:55:27 AM	85	16	1.5	0	liner
25	End	2	Peconic Bay	41.03925839	-72.18346593	2019-08-14	8:03:32 AM	85	14	1.5	0	liner
26	Start	2	Peconic Bay	NK	NK	2019-08-14	NK	85	15	1.5	0	liner

26	End	2	Peconic Bay	Nk	NK	2019-08-14	NK	85	15	1.5	0	liner
27	Start	test	RP-20	40.98926034	-72.69939738	2019-08-14	11:14:31 AM	87	0	1.5	0	liner
27	End	test	RP-20	40.98843506	-72.70290731	2019-08-14	11:26:11 AM	87	0	1.5	0	liner
28	Start	test	RP-20	40.98911282	-72.69951196	2019-08-14	11:44:00 AM	87	19	1.5	1	liner
28	End	test	RP-20	40.9887182	-72.70327603	2019-08-14	11:57:27 AM	87	0	1.5	1	liner
29	Start	test	RP-20	40.98900922	-72.69901567	2019-08-14	12:24:09 PM	87	19	1.5	0	liner
29	End	test	RP-20	40.98868442	-72.70324786	2019-08-14	12:40:57 PM	87	0	1.5	0	liner
30	Start	test	RP-20	40.98912061	-72.69914291	2019-08-14	1:12:54 PM	87	18	1.5	1	liner
30	End	test	RP-20	40.98865642	-72.7031715	2019-08-14	1:22:10 PM	87	19	1.5	1	liner
31	Start	test	RP-20	40.98941172	-72.7003085	2019-08-14	3:15:41 PM	87	15	1.5	0	winch broke, removed liner
31	End	test	RP-20	40.98971221	-72.69849809	2019-08-14	3:20:19 PM	87	15	1.5	0	no liner
32	Start	test	RP-20	40.98888986	-72.69917191	2019-08-14	3:45:49 PM	87	15	1.5	0	no liner
32	End	test	RP-20	40.98806256	-72.69564707	2019-08-14	3:52:52 PM	87	16	1.5	0	no liner
33	Start	test	RP-20	41.00412749	-72.70153628	2019-08-14	4:34:04 PM	88	18	1.5	0	no liner
33	End	test	RP-20	41.00413495	-72.69788872	2019-08-14	4:41:49 PM	87	15	1.5	0	no liner
34	Start	2	Peconic Bay	41.03802759	-72.1911352	2019-08-15	6:59:55 AM	87	17	1.5	0	no liner
34	End	2	Peconic Bay	41.03630888	-72.19222854	2019-08-15	7:15:59 AM	87	15	1.5	0	no liner
35	Start	2	Peconic Bay	41.03596011	-72.19446802	2019-08-15	7:20:28 AM	86	21	1.5	0	no liner
35	End	2	Peconic Bay	41.03687156	-72.19256574	2019-08-15	7:39:46 AM	86	22	1.5	0	no liner
36	Start	2	Peconic Bay	41.03552618	-72.19203877	2019-08-15	7:44:01 AM	86	9	1.5	0	no liner
36	End	2	Peconic Bay	41.03675949	-72.19202033	2019-08-15	7:57:39 AM	86	20	1.5	0	no liner
37	Start	2	Peconic Bay	41.03560531	-72.19344534	2019-08-15	8:01:25 AM	86	16	1.5	0	no liner
37	End	2	Peconic Bay	41.03618165	-72.19242082	2019-08-15	8:29:39 AM	32	0	1.5	0	no liner
38	Start	2	Peconic Bay	41.03511413	-72.1942086	2019-08-15	8:32:46 AM	85	13	1.5	0	no liner
38	End	2	Peconic Bay	41.0392003	-72.18388385	2019-08-15	9:04:15 AM	86	14	1.5	0	no liner
39	Start	2	Peconic Bay	41.03766491	-72.18694031	2019-08-15	9:13:04 AM	85	9	1.5	2	no liner
39	End	2	Peconic Bay	41.03990564	-72.18367162	2019-08-15	9:25:46 AM	85	0	1.5	2	no liner

40	Start	2	Peconic Bay	41.03820805	-72.18668953	2019-08-15	9:33:01 AM	86	10	1.5	0	no liner
40	End	2	Peconic Bay	41.03978394	-72.18326367	2019-08-15	9:43:19 AM	86	16	1.5	0	no liner
41	Start	2	Peconic Bay	41.03823831	-72.1863001	2019-08-15	9:51:03 AM	85	10	1.5	0	no liner
41	End	2	Peconic Bay	41.03751269	-72.18768965	2019-08-15	9:54:00 AM	85	10	1.5	0	no liner
42	Start	2	Peconic Bay	41.03927247	-72.18340776	2019-08-15	10:04:59 AM	85	14	1.5	1	no liner
42	End	2	Peconic Bay	41.0371941	-72.1874128	2019-08-15	10:15:03 AM	85	8	1.5	1	no liner
43	Start	2	Peconic Bay	41.03943458	-72.18368428	2019-08-15	10:28:39 AM	85	15	1.5	0	no liner
43	End	2	Peconic Bay	41.03781327	-72.187461	2019-08-15	10:37:20 AM	85	10	1.5	0	no liner
44	Start	3	Mecox Bay	40.88949189	-72.32573105	2019-08-15	3:47:53 PM	85	0	1.5	2	no liner
44	End	3	Mecox Bay	40.88795364	-72.32883361	2019-08-15	3:59:17 PM	85	15	1.5	2	no liner
45	Start	3	Mecox Bay	40.88937018	-72.32362744	2019-08-15	4:12:21 PM	85	16	1.5	1	no liner
45	End	3	Mecox Bay	40.88945794	-72.32481197	2019-08-15	4:20:25 PM	85	0	1.5	1	no liner
46	Start	3	Mecox Bay	40.88773009	-72.32983701	2019-08-15	4:29:21 PM	84	15	1.5	10	no liner
46	End	3	Mecox Bay	40.88777376	-72.32896362	2019-08-15	4:40:39 PM	85	19	1.5	10	no liner
47	Start	3	Mecox Bay	40.89035464	-72.32175535	2019-08-15	5:00:01 PM	84	16	1.5	2	no liner
47	End	3	Mecox Bay	40.88980872	-72.32318999	2019-08-15	5:07:31 PM	84	16	1.5	2	no liner
48	Start	3	Mecox Bay	40.8873281	-72.33073773	2019-08-16	7:48:59 AM	85	17	1.5	14	no liner
48	End	3	Mecox Bay	40.89100063	-72.32081967	2019-08-16	8:13:40 AM	84	16	1.5	14	liner
49	Start	3	Mecox Bay	40.8901507	-72.32255464	2019-08-16	8:29:59 AM	84	18	1.5	12	liner
49	End	3	Mecox Bay	40.88780452	-72.32917107	2019-08-16	8:45:10 AM	84	18	1.5	12	liner
50	Start	3	Mecox Bay	40.88736456	-72.3297438	2019-08-16	8:58:52 AM	84	21	1.5	6	liner
50	End	3	Mecox Bay	40.88920078	-72.32256311	2019-08-16	9:17:11 AM	84	22	1.5	6	liner
51	Start	3	Mecox Bay	40.88973764	-72.32265959	2019-08-16	9:31:55 AM	84	19	1.5	5	liner
51	End	3	Mecox Bay	40.88757821	-72.3298293	2019-08-16	9:47:04 AM	84	17	1.5	5	liner
52	Start	3	Mecox Bay	40.88701855	-72.33027287	2019-08-16	10:04:27 AM	84	0	1.5	12	liner
52	End	3	Mecox Bay	40.8892478	-72.32346039	2019-08-16	10:20:55 AM	84	18	1.5	12	liner
53	Start	3	Mecox Bay	40.88937001	-72.32306628	2019-08-16	10:36:34 AM	84	20	1.5	9	liner

53	End	3	Mecox Bay	40.88806302	-72.32830572	2019-08-16	10:47:01 AM	84	17	1.5	9	liner
54	Start	3	Mecox Bay	40.88766471	-72.32944448	2019-08-16	11:08:27 AM	85	18	1.5	9	liner
54	End	3	Mecox Bay	40.8886521	-72.32389013	2019-08-16	11:22:38 AM	84	20	1.5	9	liner
55	Start	3	Mecox Bay	40.88689308	-72.33011403	2019-08-16	11:34:50 AM	84	21	1.5	2	liner
55	End	3	Mecox Bay	40.88841154	-72.32478314	2019-08-16	11:47:07 AM	84	20	1.5	2	liner
56	Start	4	Shinnicock	40.83747203	-72.47068643	2019-08-16	1:10:35 PM	85	22	1.5	3	liner
56	End	4	Shinnicock	40.83852153	-72.46434518	2019-08-16	1:27:34 PM	85	24	1.5	3	liner
57	Start	4	Shinnicock	40.8364502	-72.47273429	2019-08-16	1:45:48 PM	85	20	1.5	26	liner
57	End	4	Shinnicock	40.83114454	-72.47389921	2019-08-16	1:58:08 PM	85	21	1.5	26	liner
58	Start	4	Shinnicock	40.83652002	-72.47192619	2019-08-16	2:11:51 PM	85	22	1.5	3	liner
58	End	4	Shinnicock	40.8317446	-72.47302707	2019-08-16	2:23:23 PM	85	21	1.5	3	liner
59	Start	4	Shinnicock	40.83591267	-72.47362093	2019-08-16	4:54:14 PM	86	15	1.5	5	liner
59	End	4	Shinnicock	40.83185222	-72.47314459	2019-08-16	5:04:43 PM	85	24	1.5	5	liner
60	Start	4	Shinnicock	40.83654265	-72.4730124	2019-08-16	5:46:36 PM	85	17	1.5	14	liner
60	End	4	Shinnicock	40.83194048	-72.4738133	2019-08-16	5:58:09 PM	85	0	1.5	14	liner
61	Start	4	Shinnicock	40.83171685	-72.48788022	2019-08-17	6:52:05 AM	86	15	1.5	9	liner
61	End	4	Shinnicock	40.83046979	-72.48262335	2019-08-17	7:06:30 AM	86	17	1.5	9	liner
62	Start	4	Shinnicock	40.83146732	-72.48280691	2019-08-17	7:16:22 AM	85	14	1.5	10	liner
62	End	4	Shinnicock	40.82981483	-72.4846692	2019-08-17	7:22:53 AM	85	23	1.5	10	liner
63	Start	4	Shinnicock	40.8317974	-72.48218673	2019-08-17	7:40:35 AM	85	10	1.5	3	liner
63	End	4	Shinnicock	40.82972112	-72.48418406	2019-08-17	7:51:02 AM	85	23	1.5	3	liner
64	Start	4	Shinnicock	40.83194048	-72.48681388	2019-08-17	8:04:34 AM	85	13	1.5	3	liner
64	End	4	Shinnicock	40.82992991	-72.48166597	2019-08-17	8:18:57 AM	85	20	1.5	3	liner
65	Start	4	Shinnicock	40.83218423	-72.48793052	2019-08-17	8:32:13 AM	85	11	1.5	5	liner
65	End	4	Shinnicock	40.8307987	-72.48247993	2019-08-17	8:48:40 AM	85	16	1.5	5	liner
66	Start	5	Bellport	40.71723313	-72.8979728	2019-08-17	1:13:44 PM	87	21	1.5	5	liner
66	End	5	Bellport	40.71852964	-72.89252574	2019-08-17	1:26:22 PM	87	20	1.5	5	liner

67	Start	5	Bellport	40.7174001	-72.90029291	2019-08-17	1:36:50 PM	87	16	1.5	6	liner
67	End	5	Bellport	40.71837483	-72.89644453	2019-08-17	1:56:04 PM	87	0	1.5	6	liner
68	Start	5	Bellport	40.71695443	-72.90115289	2019-08-17	2:05:34 PM	88	18	1.5	8	liner
68	End	5	Bellport	40.7183558	-72.89474459	2019-08-17	2:31:23 PM	89	18	1.5	8	liner
69	Start	5	Bellport	40.71662896	-72.90359714	2019-08-17	2:41:59 PM	88	15	1.5	7	liner
69	End	5	Bellport	40.71860039	-72.89355998	2019-08-17	3:01:35 PM	89	19	1.5	7	liner
70	Start	5	Bellport	40.72021315	-72.88984261	2019-08-17	3:17:14 PM	89	15	1.5	12	liner
70	End	5	Bellport	40.72258883	-72.88314555	2019-08-17	3:40:38 PM	87	15	1.5	12	liner
71	Start	5	Bellport	40.72223486	-72.8839399	2019-08-17	3:52:45 PM	87	14	1.5	9	liner
71	End	5	Bellport	40.71731494	-72.90110076	2019-08-17	4:39:38 PM	89	16	1.5	9	liner
72	Start	5	Bellport	40.71646962	-72.9066106	2019-08-17	5:01:39 PM	88	0	1.5	2	liner
72	End	5	Bellport	40.71602094	-72.90691168	2019-08-17	5:06:46 PM	88	12	1.5	2	liner
73	Start	5	Bellport	40.71658965	-72.90067538	2019-08-17	5:28:21 PM	89	21	1.5	5	liner
73	End	5	Bellport	40.71687036	-72.90315592	2019-08-17	5:38:26 PM	89	15	1.5	5	liner
74	Start	5	Bellport	40.71816059	-72.89698508	2019-08-17	6:01:09 PM	88	0	1.5	2	liner
74	End	5	Bellport	40.71900087	-72.89442784	2019-08-17	6:05:06 PM	88	15	1.5	2	liner
75	Start	5	Bellport	40.72171636	-72.88916107	2019-08-17	6:28:04 PM	87	13	1.5	9	liner
76	Start	6	Cupsogue	40.7552868	-72.76713263	2019-08-18	6:51:26 AM	88	12	1.5	3	liner
76	End	6	Cupsogue	40.75505421	-72.76157602	2019-08-18	7:06:27 AM	88	16	1.5	3	liner
77	Start	6	Cupsogue	40.75484742	-72.7675032	2019-08-18	7:16:11 AM	88	15	1.5	7	liner
77	End	6	Cupsogue	40.75666931	-72.76055099	2019-08-18	7:34:55 AM	87	11	1.5	7	liner
78	Start	6	Cupsogue	40.75665984	-72.76065342	2019-08-18	7:45:12 AM	87	10	1.5	15	liner
78	End	6	Cupsogue	40.75498614	-72.77171728	2019-08-18	8:16:01 AM	87	13	1.5	15	liner
79	Start	6	Cupsogue	40.75660092	-72.76051001	2019-08-18	8:29:06 AM	87	11	1.5	13	liner
79	End	6	Cupsogue	40.75579525	-72.7717208	2019-08-18	8:56:30 AM	86	14	1.5	13	liner
80	Start	6	Cupsogue	40.75647594	-72.77044055	2019-08-18	9:07:44 AM	87	12	1.5	3	liner
80	End	6	Cupsogue	40.75725261	-72.7613834	2019-08-18	9:31:54 AM	86	8	1.5	3	liner

81	Start	6	Cupsogue	40.75867242	-72.75241912	2019-08-18	9:44:47 AM	86	14	1.5	8	liner
81	End	6	Cupsogue	40.76280395	-72.74887743	2019-08-18	10:04:40 AM	86	0	1.5	8	liner
82	Start	6	Cupsogue	40.75827654	-72.75283051	2019-08-18	10:12:11 AM	86	13	1.5	11	liner
82	End	6	Cupsogue	40.76354306	-72.74409304	2019-08-18	10:32:32 AM	87	16	1.5	11	liner
83	Start	6	Cupsogue	40.75823028	-72.75216071	2019-08-18	10:46:05 AM	86	16	1.5	11	liner
83	End	6	Cupsogue	40.76379712	-72.74358158	2019-08-18	11:11:45 AM	87	15	1.5	11	liner
84	Start	6	Cupsogue	40.75392533	-72.76999296	2019-08-18	2:03:29 PM	88	19	1.5	17	liner
84	End	6	Cupsogue	40.75404653	-72.76034128	2019-08-18	2:20:34 PM	88	26	1.5	17	liner
85	Start	6	Cupsogue	40.75453981	-72.7602018	2019-08-18	2:28:56 PM	88	22	1.5	15	liner
85	End	6	Cupsogue	40.75468029	-72.76098702	2019-08-18	2:30:30 PM	88	17	1.5	15	liner
86	Start	6	Cupsogue	40.75345988	-72.76904211	2019-08-18	2:51:48 PM	89	19	1.5	30	liner
86	End	6	Cupsogue	40.75343356	-72.76970328	2019-08-18	3:08:02 PM	89	19	1.5	30	liner
87	Start	6	Cupsogue	40.7555344	-72.76381147	2019-08-18	3:14:09 PM	88	0	1.5	16	liner
87	End	6	Cupsogue	40.75353255	-72.76176193	2019-08-18	3:26:21 PM	88	27	1.5	16	liner
88	Start	6	Cupsogue	40.75771554	-72.75211561	2019-08-18	3:42:08 PM	89	20	1.5	6	liner
88	End	6	Cupsogue	40.75771521	-72.75211411	2019-08-18	3:42:08 PM	89	20	1.5	6	liner
89	Start	6	Cupsogue	40.76250656	-72.74309468	2019-08-18	4:00:02 PM	89	18	1.5	2	liner
89	End	6	Cupsogue	40.76211395	-72.74361678	2019-08-18	4:11:57 PM	88	19	1.5	2	liner
90	Start	6	Cupsogue	40.75850395	-72.74898338	2019-08-18	4:35:39 PM	89	25	1.5	5	liner
90	End	6	Cupsogue	40.75860905	-72.75090795	2019-08-18	4:46:33 PM	90	18	1.5	5	liner

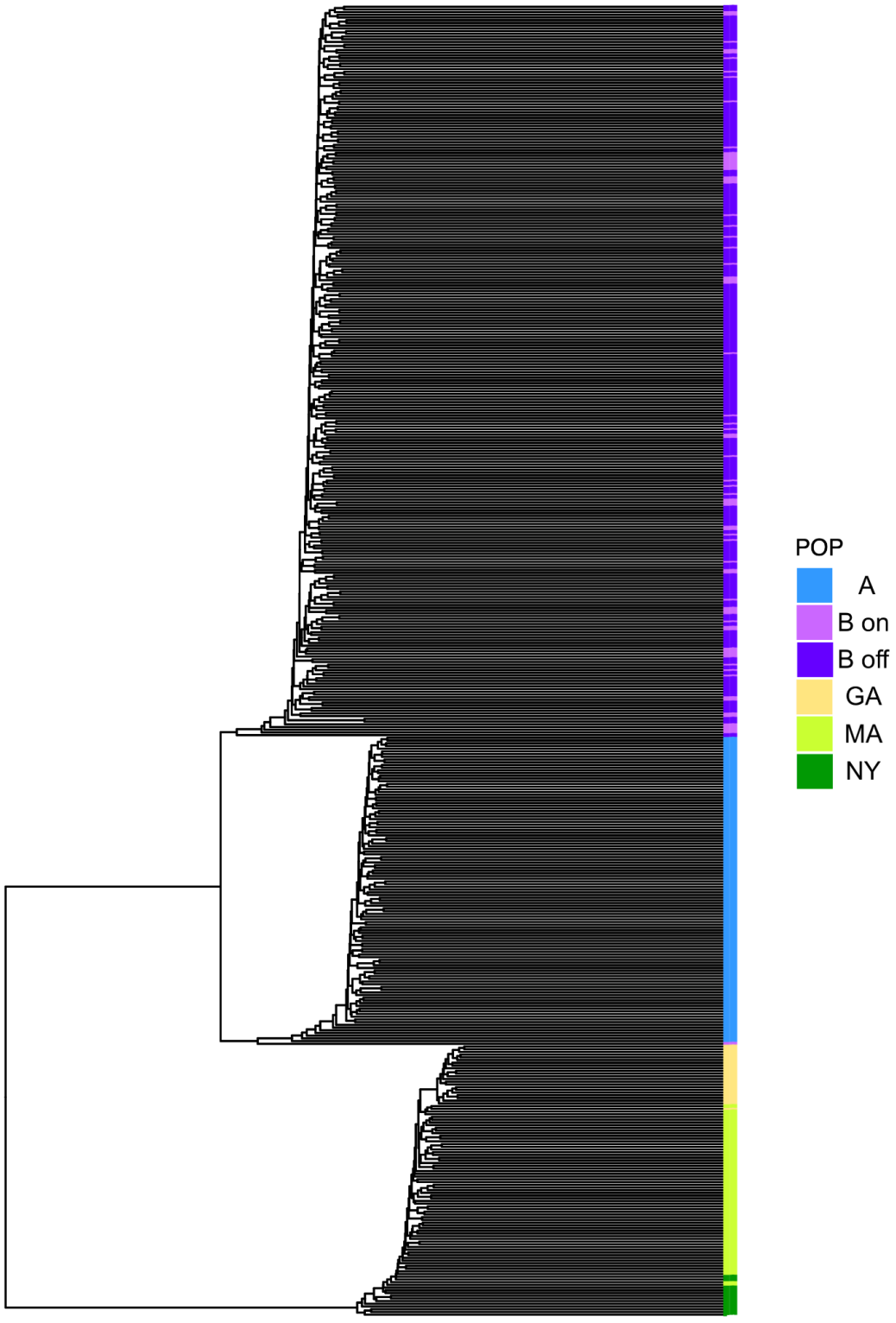

Figure A.3: Phylogeny of individual-level UPGMA phylogenetic tree for 483 Spisula solidissima sp. individuals based calculated in Vcf2PopTree (from 4.7 LD pruned SNPs, including all OTUs and hybrids) and plotted with ggtree R package. Branch lengths are scaled to represent the percentage of genetic variation.

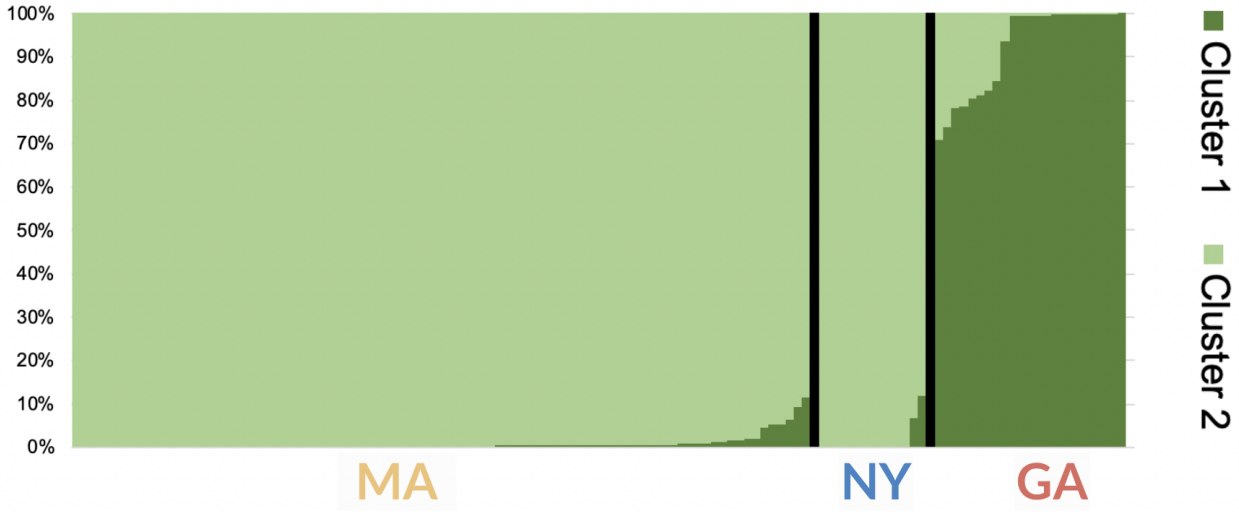

Figure A.4: Clustering and admixture results from program STRUCTURE applied to all S.s. similis samples using 1.2 thousand haplotype loci. Models assuming K=2 source populations showed the greatest support from the data, here depicted as dark and light green. Black vertical lines separate sampling regions geographically (MA = Massachusetts, NY = New York, GA = Georgia). Each individual specimen is represented with a thin vertical bar, where a combination of the two greens indicates proportional contributions from the two cluster sources (admixture). Individual surfclams are arbitrarily ordered from fully cluster 2 to increasing proportions of cluster 1.

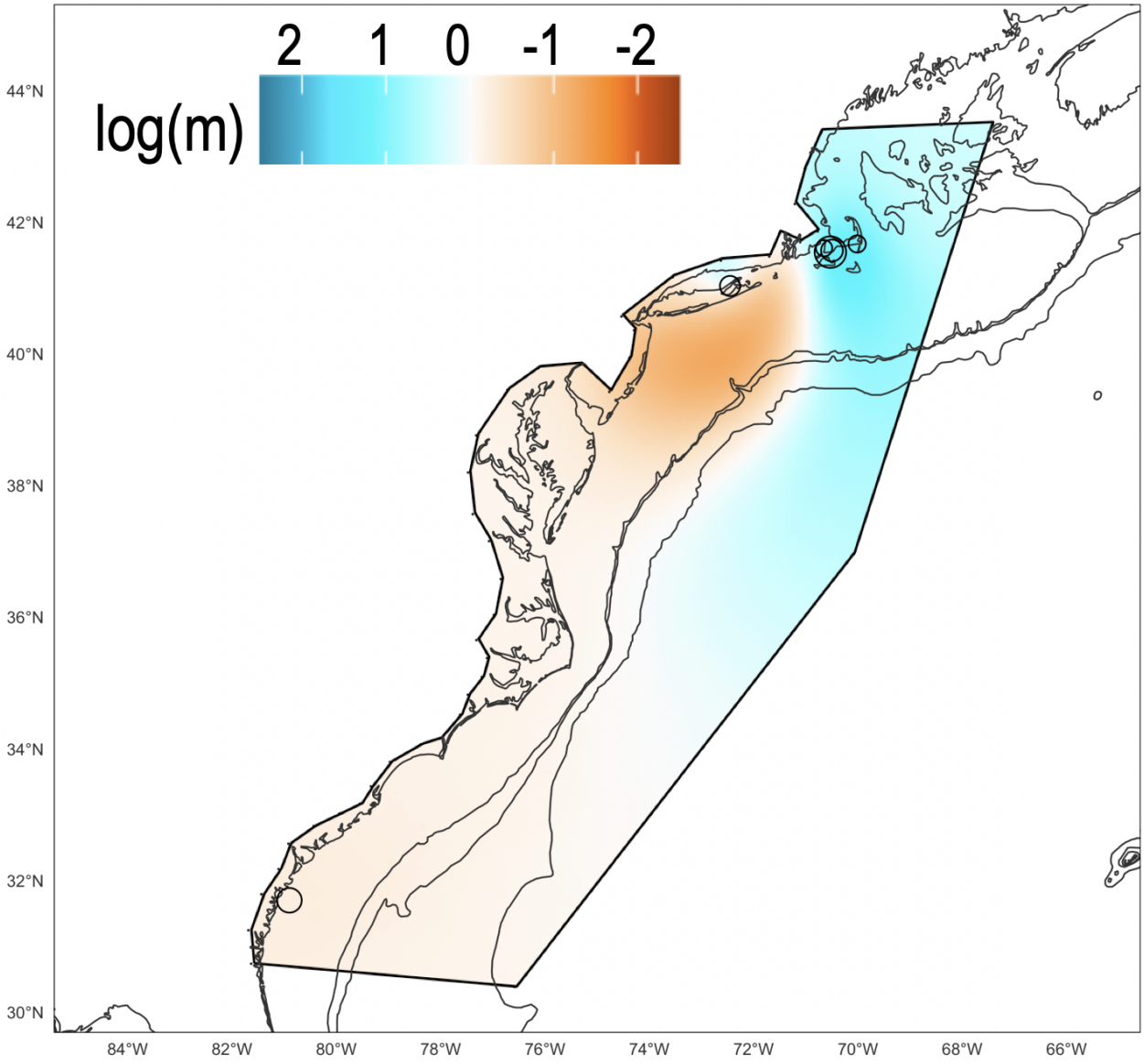

Figure A.5: S.s. similis EEMS results. EEMS estimates of connectivity are illustrated relative to the null hypothesis of isolation by distance shown in white (log(m) = 0). Darker orange areas show estimated increased genetic differentiation whereas darker blue areas represent areas of greater genetic similarity. The black polygon describes the area in which the EEMS analysis was conducted. Circles show sampling locations (from Fig. 1) scaled by the log of the number of samples collected at each location.

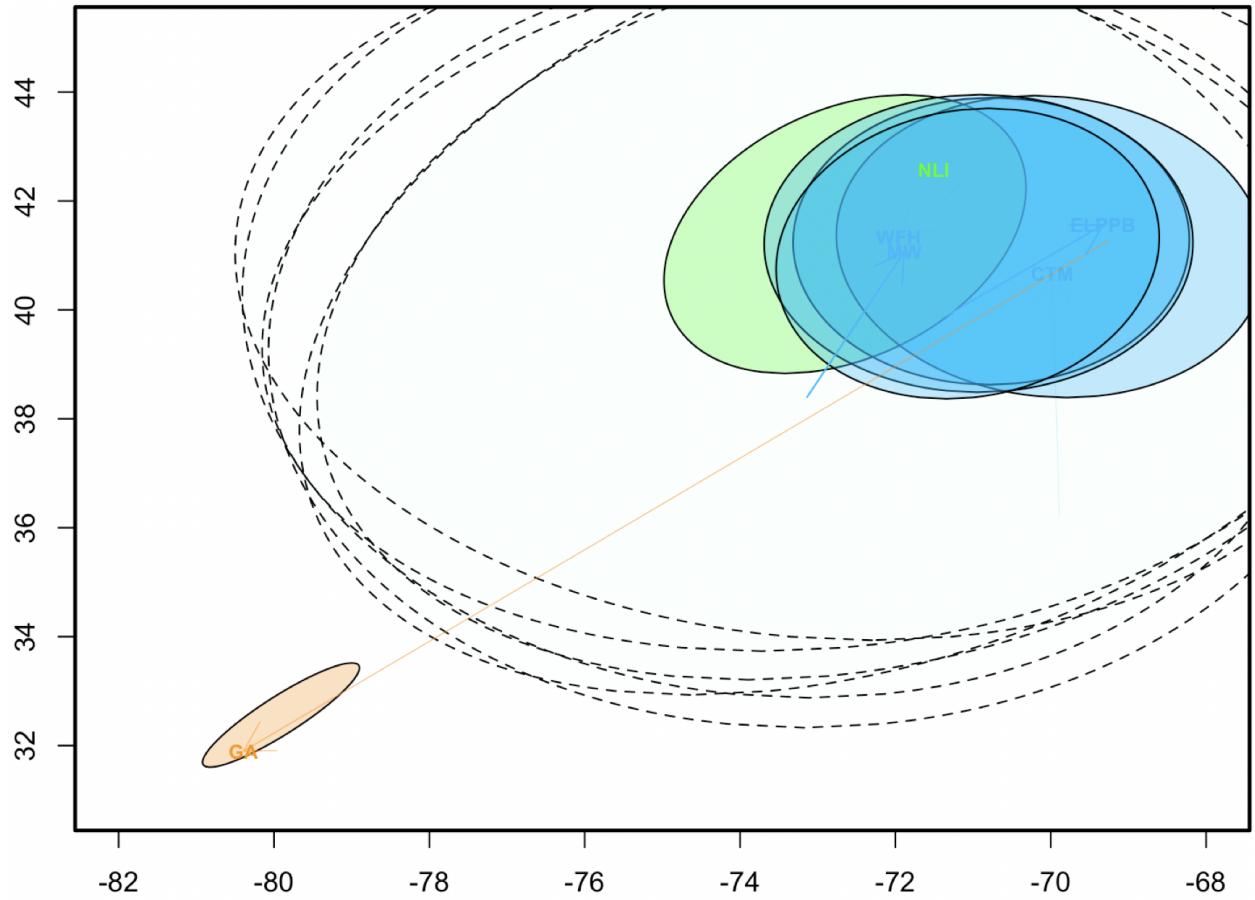

Figure A.6: S.s. similis SpaceMix results. Georgia populations are shown in orange while Long Island, NY populations are shown in green and populations from the Southern coast of Massachusetts are shown in blue. Spacemix results for populations, colored by sampling region (Table A.1), are shown in "geogenetic space" which combines geographic sampling location inputs with genetic distance information. The axes show an adjusted "latitude" and "longitude" in geogenetic space rather than actual cartesian coordinates. The colored text labels show the geogenetic location estimate of each population, with the corresponding colored ellipses showing 95% confidence intervals on those estimates. Arrows depict the estimated direction and relative magnitude of admixture in the population. Arrows originate at the estimated origin of the detected admixture, pointing towards the colored label at the population location estimate. Black dotted ellipses show 95% confidence intervals for the origin point of the admixture arrows.

In Fig. A.6, we observe that the origin for the admixture in Georgia (GA, orange) stems from the geogenetic space inhabited by the Massachusetts populations (blue) with the 95% confidence interval surrounding the Massachusetts sites but not the estimate of the Georgia population. This is the best indicator we have of directional gene flow from north to south in *S.s. similis*.

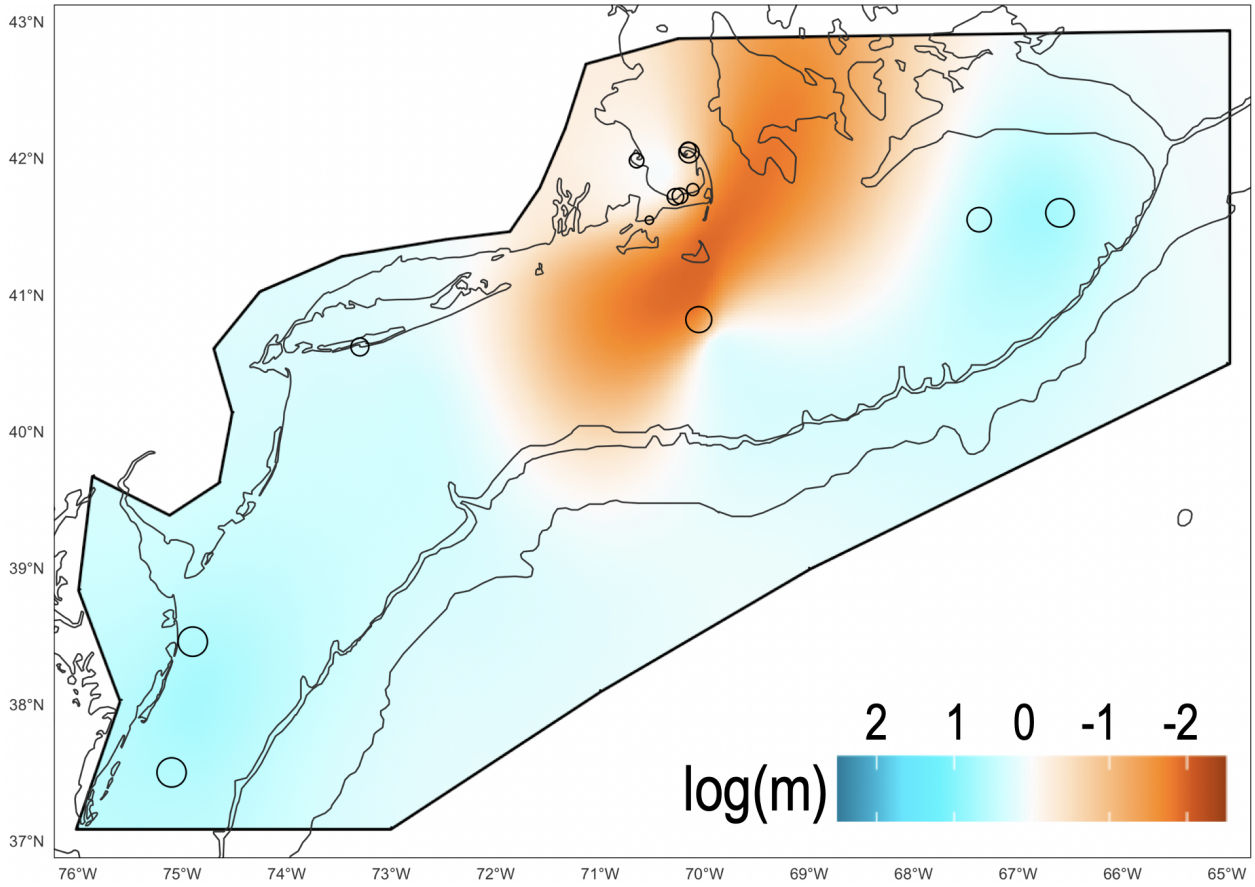

Figure A.7: S.s. solidissima OTU B EEMS results. EEMS estimates of connectivity are illustrated relative to the null hypothesis of isolation by distance shown in white (log(m) = 0). Darker orange areas show estimated increased genetic differentiation whereas darker blue areas represent areas of greater genetic similarity. The black polygon describes the area in which the EEMS analysis was conducted. Circles show sampling locations (from Fig. 1) scaled by the log of the number of samples collected at each location.

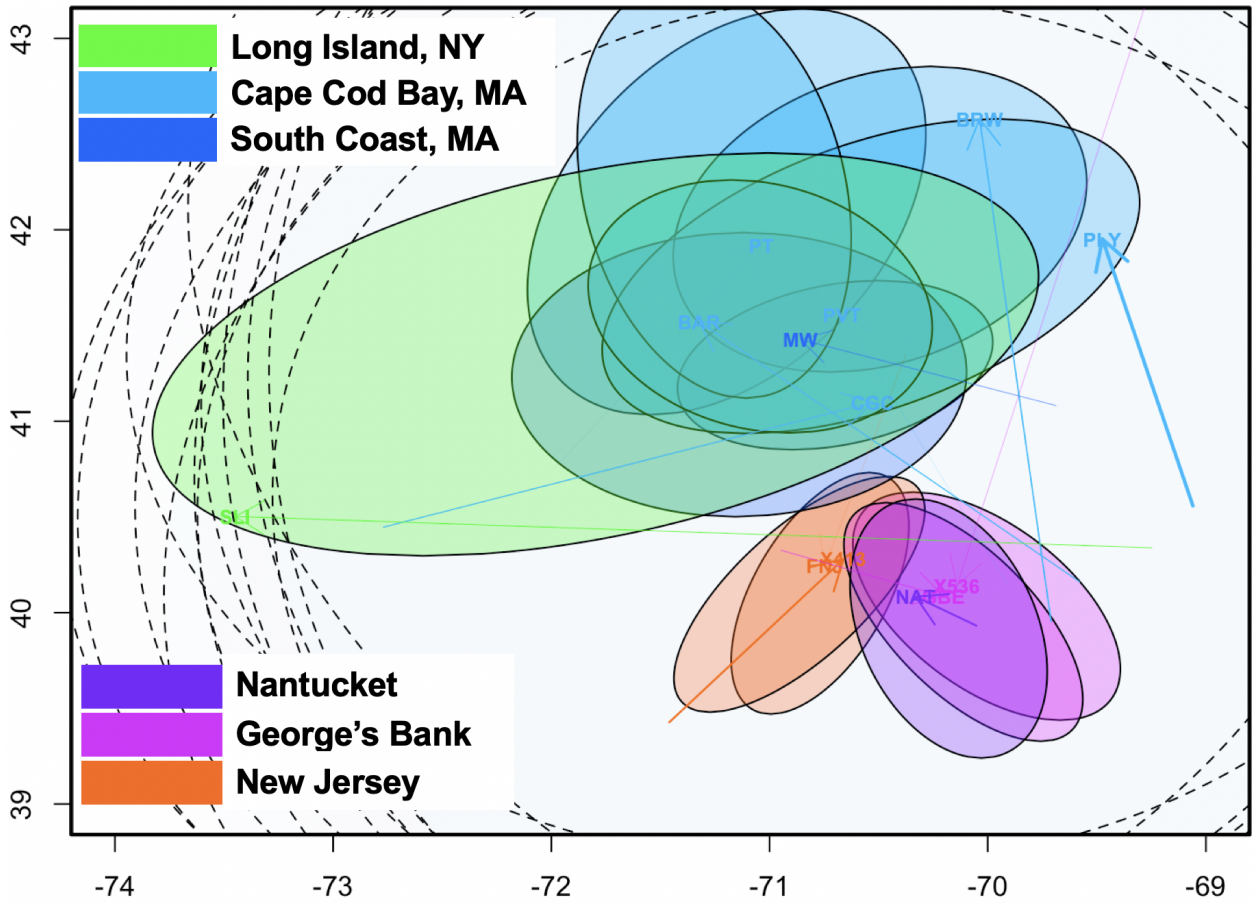

Figure A.8: Spacemix results for S.s. solidissima OTU B populations, colored by sampling region (Table A.1), are shown in "geogenetic space" which combines geographic sampling location inputs with genetic distance information. The axes show an adjusted "latitude" and "longitude" in geogenetic space rather than actual cartesian coordinates. The colored text labels show the geogenetic location estimate of each population, with the corresponding colored ellipses showing 95% confidence intervals on those estimates. Arrows depict the estimated direction and relative magnitude of admixture in the population. Arrows originate at the estimated origin of the detected admixture, pointing towards the colored label at the population location estimate. Black dotted ellipses show 95% confidence intervals for the origin point of the admixture arrows.

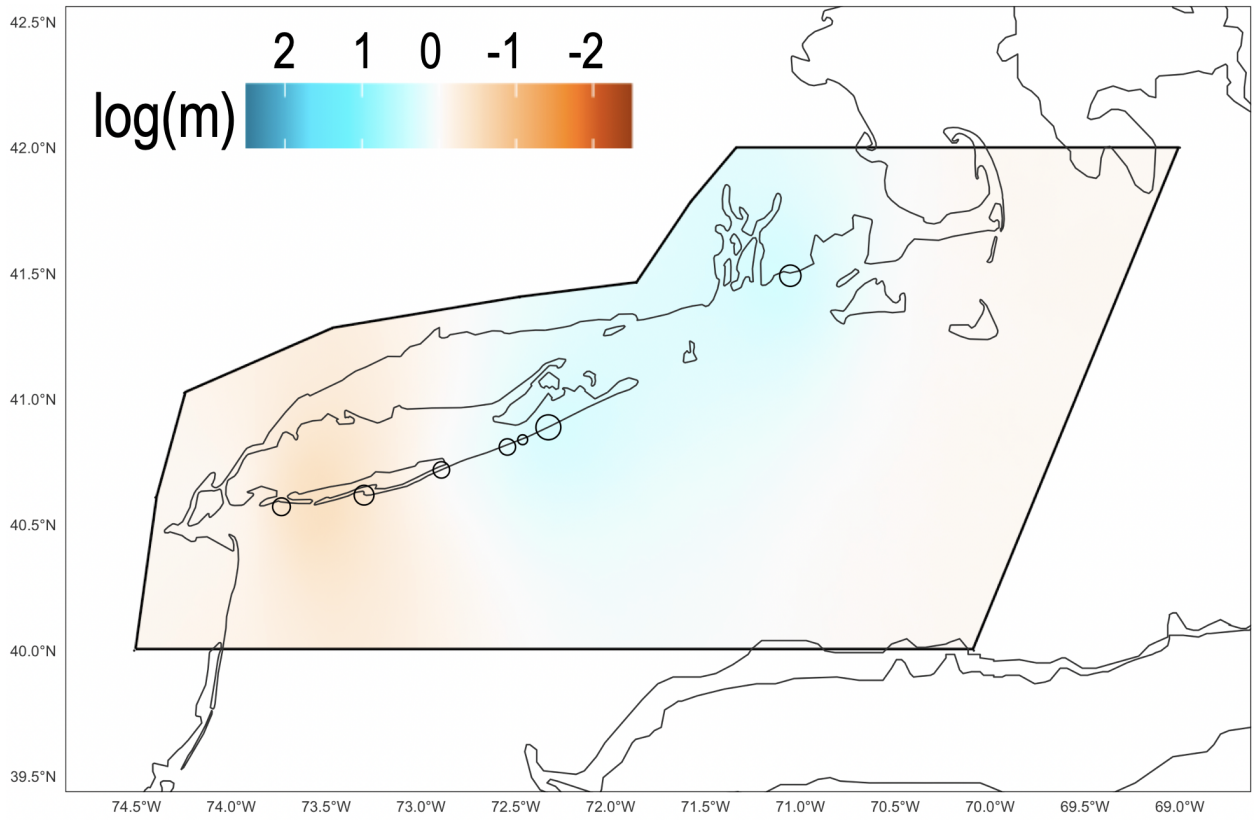

Figure A.9: S.s. solidissima OTU A EEMS results. EEMS estimates of connectivity are illustrated relative to the null hypothesis of isolation by distance shown in white (log(m) = 0). Darker orange areas show estimated increased genetic differentiation whereas darker blue areas represent areas of greater genetic similarity. The black polygon describes the area in which the EEMS analysis was conducted. Circles show sampling locations (from Fig. 1) scaled by the log of the number of samples collected at each location.

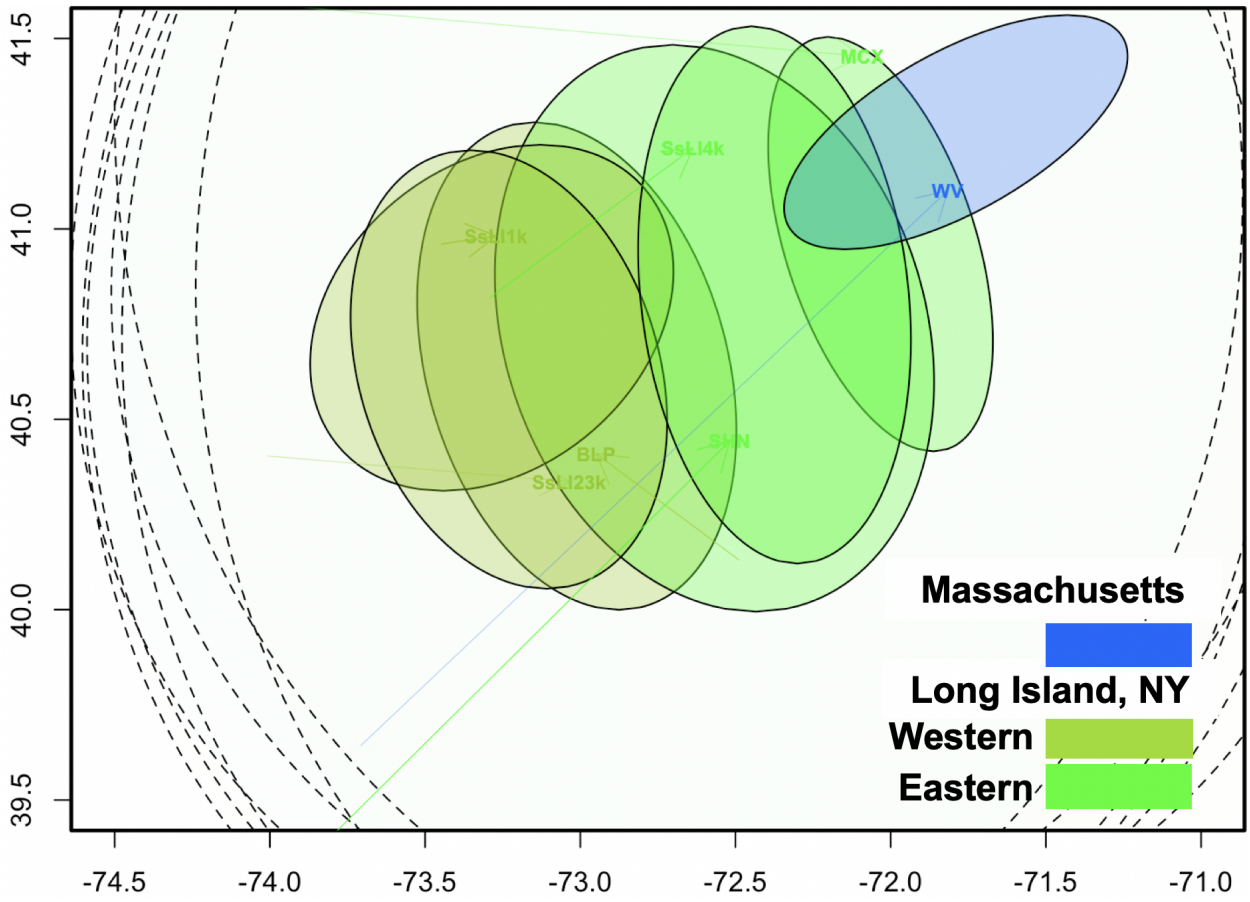

Figure A.10: Spacemix results for S.s. solidissima OTU A populations, colored by sampling region (Table A.1), are shown in "geogenetic space" which combines geographic sampling location inputs with genetic distance information. The axes show an adjusted "latitude" and "longitude" in geogenetic space rather than actual cartesian coordinates. The colored text labels show the geogenetic location estimate of each population, with the corresponding colored ellipses showing 95% confidence intervals on those estimates. Arrows depict the estimated direction and relative magnitude of admixture in the population. Arrows originate at the estimated origin of the detected admixture, pointing towards the colored label at the population location estimate. Black dotted ellipses show 95% confidence intervals for the origin point of the admixture arrows.

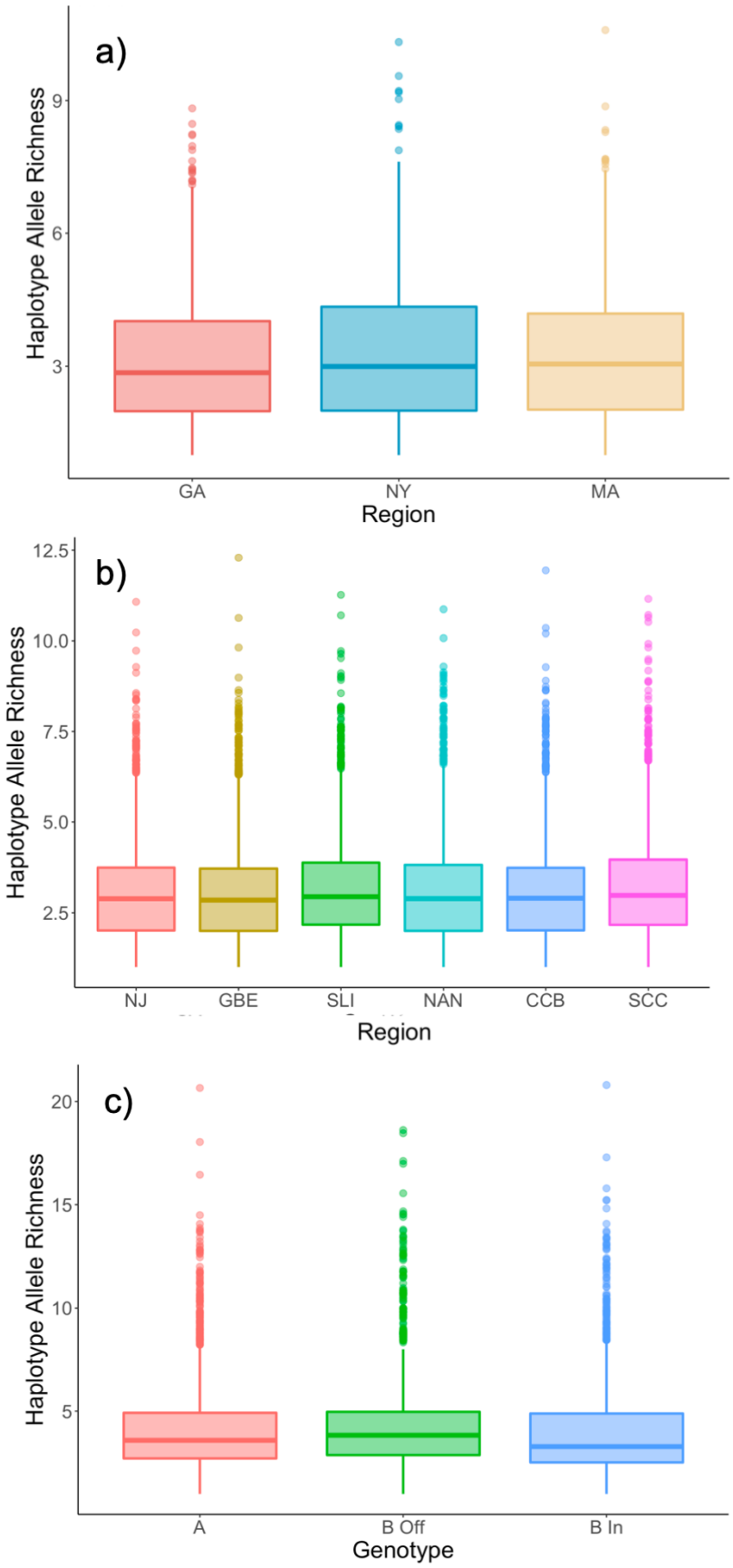

Figure A. 11: Distributions of haplotype allele richness across sampling regions of (a) S.s. similis, and (b) S.s. solidissima, and (c) OTUs of S.s. solidissima. Outlier points are shown as dots, while the box plot represents the third and first quartile with the center line indicating the mean value. (a) GA = Georgia, NY = New York, MA = Massachusetts. (b) NJ = New Jersey, GBE = George's Bank, SLI = Southern Long Island, NY, NAN = Nantucket, CCB = Cape Cod Bay, MA, SCC = Southern Coast of Cape Cod, MA. (c) A = OTU A, B Off = Samples from OTU B found at offshore sites, B In = Samples from OTU B found at inshore sites.

Table A.3: Bioinformatic data sets and the analyses each was used for.

Data Set	**Purpose**	**Analyses Used In**
Basic SNP data	After aligning samples to dDocent created catalog, we followed Puritz's recommended SNP filtering protocol (http://www.ddocent.com/filtering/)	None, further filter steps were performed as outlined below
LD (r^2<0.8) pruned	Linkage Disequilibrium (LD) is the correlation between nearby variants such that the alleles at neighboring polymorphisms are associated within a population more often than if they were unlinked. Most analyses must act under the assumption that loci used are independent of each other. r^2 (the squared correlation based on genotypic allele counts) set to less than 0.8 is a commonly used method for reducing correlations between loci in ddRAD data with no reference genome. After filtering the linkage between dDocent catalog contigs, SNPs were also filtered to one per contig.	Subspecies-specific PCA, F_{ST} and AMOVA, EEMS, SpaceMix, SNAPP
LD (r^2<0.2) pruned	Pruning to include only loci with r^2<0.2 is a more stringent filter against correlated loci used for analyses which are particularly sensitive to linked loci.	UPGMA tree
Haplotype Data	Sets of closely linked SNPs (from the basic SNP data) inferred to be a haplotype using rad_haplotyper. For some analyses, this gives us more statistical power by giving each maker more than two possible values (unlike SNPs for which there are only two values at each locus.	STRUCTURE
PCA Subset	For the original outline of the difference between the OTUs, a subset of *S.s. similis* were used to avoid batch effects from sequencing; only those sequenced with *S.s. solidissima* were used.	Both subspecies PCA
N_e Subset	NeEstimator is incredibly computationally expensive at high numbers of loci so SNP loci were subset randomly from the LD (r^2<0.8) pruned data to 1260 SNPs for *S.s. similis*, 2540 SNPs for *S.s. solidissima*	N_e Estimation
Diversity Subset	Because diversity estimates, even with computational corrections, are sensitive to sample and marker number, we subset both *S.s. solidissima* and *S.s. similis* to 13 individuals per OTU/region and 1050 haplotype loci.	Heterozygosity, Haplotype Allele Richness

Table A.4: Sample information from 2019 tows on Long Island by Matt Weeks. OTU designation is given if the sample was sequenced, but subspecies-level information is provided for all samples from RFLP (restriction fragment length polymorphism) gel electrophoresis tests based on mitochondrial markers from Hare & Borchardt-Wier (2014).

Sample ID	Station	Site ID	Subspecies	OTU	Tow	Length (cm)
1	1	GLD	*S.s. similis*	similis	2	3.2
2	1	GLD	*S.s. similis*	similis	7	
3	1	GLD	*S.s. similis*	similis	13	3.6
4	1	GLD	*S.s. similis*	similis	13	8.3
5	1	GLD	*S.s. similis*	similis	17	4.7
6	1	GLD	*S.s. similis*	similis	17	4
7	1	GLD	*S.s. similis*	similis	20	2.5
8	1	GLD	*S.s. similis*	similis	22	3
9	test	RP20	*S.s. similis*	similis	28	2.1
10	test	RP20	*S.s. similis*	similis	30	1
11	2	PEC	*S.s. similis*	similis	39	5
12	2	PEC	*S.s. similis*	similis	39	5.4
13	2	MCX	*S.s. similis*	similis	42	6.5
14	3	MCX	*S.s. solidissima*	unsequenced	44	6.5
15	3	MCX	*S.s. solidissima*	A	44	6.5
16	3	MCX	*S.s. solidissima*	A	45	6.8
17	3	MCX	*S.s. solidissima*	A	46	8.3
18	3	MCX	*S.s. solidissima*	A	46	6.5
19	3	MCX	*S.s. solidissima*	A	46	6.9
20	3	MCX	*S.s. solidissima*	A	46	7.6
21	3	MCX	*S.s. solidissima*	A	46	7.6
22	3	MCX	*S.s. solidissima*	unsequenced	46	5.7
23	3	MCX	*S.s. solidissima*	unsequenced	46	9.1
24	3	MCX	*S.s. solidissima*	unsequenced	46	7
25	3	MCX	*S.s. solidissima*	unsequenced	46	
26	3	MCX	*S.s. solidissima*	unsequenced	46	
27	3	MCX	*S.s. solidissima*	A	47	6.4
28	3	MCX	*S.s. solidissima*	unsequenced	47	
29	3	MCX	*S.s. solidissima*	A	48	8.6
30	3	MCX	*S.s. solidissima*	A	48	7.6

31	3	MCX	*S.s. solidissima*	A	48	7.5
32	3	MCX	*S.s. solidissima*	A	48	6.5
33	3	MCX	*S.s. solidissima*	A	48	6
34	3	MCX	*S.s. solidissima*	A	48	5.9
35	3	MCX	*S.s. solidissima*	A	48	6.5
36	3	MCX	*S.s. solidissima*	unsequenced	48	5.3
37	3	MCX	*S.s. solidissima*	unsequenced	48	6.8
38	3	MCX	*S.s. solidissima*	shows admixture (90% A, 10% B)	48	7.5
39	3	MCX	*S.s. solidissima*	A	48	7.3
40	3	MCX	*S.s. solidissima*	unsequenced	48	
41	3	MCX	*S.s. solidissima*	unsequenced	48	
42	3	MCX	*S.s. solidissima*	unsequenced	48	
43	3	MCX	*S.s. solidissima*	unsequenced	49	6.3
44	3	MCX	*S.s. solidissima*	unsequenced	49	7.2
45	3	MCX	*S.s. solidissima*	unsequenced	49	4.6
46	3	MCX	*S.s. solidissima*	A	49	6.7
47	3	MCX	*S.s. solidissima*	A	49	6.6
48	3	MCX	*S.s. solidissima*	unsequenced	49	
49	3	MCX	*S.s. solidissima*	unsequenced	49	
50	3	MCX	*S.s. solidissima*	unsequenced	49	
51	3	MCX	*S.s. solidissima*	unsequenced	49	10.2
52	3	MCX	*S.s. solidissima*	unsequenced	49	
53	3	MCX	*S.s. solidissima*	unsequenced	49	
54	3	MCX	*S.s. solidissima*	unsequenced	49	
55	3	MCX	*S.s. solidissima*	A	50	6.2
56	3	MCX	*S.s. solidissima*	unsequenced	50	4
57	3	MCX	*S.s. solidissima*	unsequenced	50	4.9
58	3	MCX	*S.s. solidissima*	unsequenced	50	5.4
59	3	MCX	*S.s. solidissima*	unsequenced	50	
60	3	MCX	*S.s. solidissima*	unsequenced	50	
61	3	MCX	*S.s. solidissima*	A	51	6.6
62	3	MCX	*S.s. solidissima*	A	51	6.6
63	3	MCX	*S.s. solidissima*	unsequenced	51	8.3
64	3	MCX	*S.s. solidissima*	unsequenced	51	4.5

65	3	MCX	*S.s. solidissima*	unsequenced	51	
66	3	MCX	*S.s. solidissima*	A	52	6.2
67	3	MCX	*S.s. solidissima*	A	52	6.6
68	3	MCX	*S.s. solidissima*	A	52	6.5
69	3	MCX	*S.s. solidissima*	A	52	6
70	3	MCX	*S.s. solidissima*	unsequenced	52	6.2
71	3	MCX	*S.s. solidissima*	unsequenced	52	5.5
72	3	MCX	*S.s. solidissima*	unsequenced	52	
73	3	MCX	*S.s. solidissima*	unsequenced	52	
74	3	MCX	*S.s. solidissima*	unsequenced	52	
75	3	MCX	*S.s. solidissima*	unsequenced	52	
76	3	MCX	*S.s. solidissima*	unsequenced	52	
77	3	MCX	*S.s. solidissima*	unsequenced	52	
78	3	MCX	*S.s. solidissima*	unsequenced	53	4.9
79	3	MCX	*S.s. solidissima*	A	53	6.4
80	3	MCX	*S.s. solidissima*	A	53	6.5
81	3	MCX	*S.s. solidissima*	unsequenced	53	5.8
82	3	MCX	*S.s. solidissima*	unsequenced	53	6
83	3	MCX	*S.s. solidissima*	unsequenced	53	
84	3	MCX	*S.s. solidissima*	A	53	7.6
85	3	MCX	*S.s. solidissima*	unsequenced	53	
86	3	MCX	*S.s. solidissima*	unsequenced	53	
87	3	MCX	*S.s. solidissima*	unsequenced	54	4.4
88	3	MCX	*S.s. solidissima*	unsequenced	54	5.4
89	3	MCX	*S.s. solidissima*	A	54	5.8
90	3	MCX	*S.s. solidissima*	A	54	6
91	3	MCX	*S.s. solidissima*	A	54	7.1
92	3	MCX	*S.s. solidissima*	unsequenced	54	5.1
93	3	MCX	*S.s. solidissima*	A	54	7
94	3	MCX	*S.s. solidissima*	unsequenced	54	
95	3	MCX	*S.s. solidissima*	unsequenced	54	
96	3	MCX	*S.s. solidissima*	A	55	5.9
97	3	MCX	*S.s. solidissima*	unsequenced	55	
98	4	SHN	*S.s. solidissima*	unsequenced	56	4.4
99	4	SHN	*S.s. solidissima*	unsequenced	56	6.1

100	4	SHN	*S.s. solidissima*	unsequenced	56	
101	4	SHN	*S.s. solidissima*	unsequenced	57	6.1
102	4	SHN	*S.s. solidissima*	A	57	6.5
103	4	SHN	*S.s. solidissima*	A	57	7.5
104	4	SHN	*S.s. solidissima*	B	57	8.7
105	4	SHN	*S.s. solidissima*	unsequenced	57	5.5
106	4	SHN	*S.s. solidissima*	unsequenced	57	10.7
107	4	SHN	*S.s. solidissima*	unsequenced	57	
108	4	SHN	*S.s. solidissima*	unsequenced	57	
109	4	SHN	*S.s. solidissima*	unsequenced	57	
110	4	SHN	*S.s. solidissima*	unsequenced	57	
111	4	SHN	*S.s. solidissima*	unsequenced	57	
112	4	SHN	*S.s. solidissima*	unsequenced	57	
113	4	SHN	*S.s. solidissima*	unsequenced	57	
114	4	SHN	*S.s. solidissima*	unsequenced	57	13.5
115	4	SHN	*S.s. solidissima*	unsequenced	57	
116	4	SHN	*S.s. solidissima*	unsequenced	57	
117	4	SHN	*S.s. solidissima*	unsequenced	57	
118	4	SHN	*S.s. solidissima*	unsequenced	57	13.5
119	4	SHN	*S.s. solidissima*	unsequenced	57	
120	4	SHN	*S.s. solidissima*	unsequenced	57	
121	4	SHN	*S.s. solidissima*	unsequenced	57	
122	4	SHN	*S.s. solidissima*	unsequenced	57	
123	4	SHN	*S.s. solidissima*	unsequenced	57	13.3
124	4	SHN	*S.s. solidissima*	unsequenced	57	
125	4	SHN	*S.s. solidissima*	unsequenced	57	
126	4	SHN	*S.s. solidissima*	unsequenced	57	
127	4	SHN	*S.s. solidissima*	unsequenced	58	
128	4	SHN	*S.s. solidissima*	unsequenced	58	
129	4	SHN	*S.s. solidissima*	unsequenced	58	
130	4	SHN	*S.s. solidissima*	unsequenced	59	5.8
131	4	SHN	*S.s. solidissima*	unsequenced	59	4.4
132	4	SHN	*S.s. solidissima*	unsequenced	59	7
133	4	SHN	*S.s. solidissima*	unsequenced	59	
134	4	SHN	*S.s. solidissima*	unsequenced	59	

135	4	SHN	*S.s. solidissima*	unsequenced	60	3.9
136	4	SHN	*S.s. solidissima*	unsequenced	60	5.6
137	4	SHN	*S.s. solidissima*	unsequenced	60	5.6
138	4	SHN	*S.s. solidissima*	unsequenced	60	5.5
139	4	SHN	*S.s. solidissima*	unsequenced	60	7
140	4	SHN	*S.s. solidissima*	unsequenced	60	7.5
141	4	SHN	*S.s. solidissima*	unsequenced	60	
142	4	SHN	*S.s. solidissima*	unsequenced	60	13.5
143	4	SHN	*S.s. solidissima*	unsequenced	60	
144	4	SHN	*S.s. solidissima*	unsequenced	60	
145	4	SHN	*S.s. solidissima*	unsequenced	60	13.4
146	4	SHN	*S.s. solidissima*	unsequenced	60	
147	4	SHN	*S.s. solidissima*	unsequenced	60	
148	4	SHN	*S.s. solidissima*	unsequenced	60	
149	4	SHN	*S.s. solidissima*	unsequenced	61	4.9
150	4	SHN	*S.s. solidissima*	unsequenced	61	6.1
151	4	SHN	*S.s. solidissima*	A	61	6.7
152	4	SHN	*S.s. solidissima*	A	61	6
153	4	SHN	*S.s. solidissima*	unsequenced	61	7.6
154	4	SHN	*S.s. solidissima*	unsequenced	61	
155	4	SHN	*S.s. solidissima*	unsequenced	61	13.5
156	4	SHN	*S.s. solidissima*	unsequenced	61	
157	4	SHN	*S.s. solidissima*	unsequenced	61	
158	4	SHN	*S.s. solidissima*	unsequenced	62	5.1
159	4	SHN	*S.s. solidissima*	unsequenced	62	5.6
160	4	SHN	*S.s. solidissima*	unsequenced	62	4.5
161	4	SHN	*S.s. solidissima*	unsequenced	62	5.6
162	4	SHN	*S.s. solidissima*	unsequenced	62	
163	4	SHN	*S.s. solidissima*	unsequenced	62	
164	4	SHN	*S.s. solidissima*	unsequenced	62	12.5
165	4	SHN	*S.s. solidissima*	unsequenced	62	14
166	4	SHN	*S.s. solidissima*	unsequenced	62	12.1
167	4	SHN	*S.s. solidissima*	unsequenced	62	
168	4	SHN	*S.s. solidissima*	unsequenced	63	5
169	4	SHN	*S.s. solidissima*	unsequenced	63	6

170	4	SHN	*S.s. solidissima*	unsequenced	63	
171	4	SHN	*S.s. solidissima*	unsequenced	64	5.2
172	4	SHN	*S.s. solidissima*	unsequenced	64	5.5
173	4	SHN	*S.s. solidissima*	unsequenced	64	6.5
174	4	SHN	*S.s. solidissima*	unsequenced	65	
175	4	SHN	*S.s. solidissima*	unsequenced	65	
176	4	SHN	*S.s. solidissima*	unsequenced	65	
177	4	SHN	*S.s. solidissima*	unsequenced	65	
178	4	SHN	*S.s. solidissima*	unsequenced	65	
179	5	BLP	*S.s. solidissima*	A	66	4.4
180	5	BLP	*S.s. solidissima*	unsequenced	66	
181	5	BLP	*S.s. solidissima*	unsequenced	66	
182	5	BLP	*S.s. solidissima*	unsequenced	66	
183	5	BLP	*S.s. solidissima*	unsequenced	66	
184	5	BLP	*S.s. solidissima*	A	67	4.2
185	5	BLP	*S.s. solidissima*	unsequenced	67	
186	5	BLP	*S.s. solidissima*	unsequenced	67	
187	5	BLP	*S.s. solidissima*	unsequenced	67	
188	5	BLP	*S.s. solidissima*	unsequenced	67	
189	5	BLP	*S.s. solidissima*	unsequenced	67	
190	5	BLP	*S.s. solidissima*	A	68	6.2
191	5	BLP	*S.s. solidissima*	A	68	
192	5	BLP	*S.s. solidissima*	unsequenced	68	
193	5	BLP	*S.s. solidissima*	unsequenced	68	
194	5	BLP	*S.s. solidissima*	unsequenced	68	
195	5	BLP	*S.s. solidissima*	unsequenced	68	
196	5	BLP	*S.s. solidissima*	unsequenced	68	
197	5	BLP	*S.s. solidissima*	unsequenced	68	
198	5	BLP	*S.s. solidissima*	A	69	3.4
199	5	BLP	*S.s. solidissima*	A	69	2.9
200	5	BLP	*S.s. solidissima*	A	69	3.3
201	5	BLP	*S.s. solidissima*	A	69	2.5
202	5	BLP	*S.s. solidissima*	A	69	3.5
203	5	BLP	*S.s. solidissima*	unsequenced	69	
204	5	BLP	*S.s. solidissima*	unsequenced	69	

205	5	BLP	*S.s. solidissima*	unsequenced	70	
206	5	BLP	*S.s. solidissima*	unsequenced	70	
207	5	BLP	*S.s. solidissima*	unsequenced	70	
208	5	BLP	*S.s. solidissima*	unsequenced	70	
209	5	BLP	*S.s. solidissima*	unsequenced	70	
210	5	BLP	*S.s. solidissima*	A	70	13
211	5	BLP	*S.s. solidissima*	A	70	3
212	5	BLP	*S.s. solidissima*	A	70	6.5
213	5	BLP	*S.s. solidissima*	B	70	6.1
214	5	BLP	*S.s. solidissima*	A	70	3.5
215	5	BLP	*S.s. solidissima*	A	70	3.3
216	5	BLP	*S.s. solidissima*	A	70	2.5
217	5	BLP	*S.s. solidissima*	unsequenced	71	
218	5	BLP	*S.s. solidissima*	unsequenced	71	
219	5	BLP	*S.s. solidissima*	unsequenced	71	
220	5	BLP	*S.s. solidissima*	unsequenced	71	
221	5	BLP	*S.s. solidissima*	unsequenced	71	
222	5	BLP	*S.s. solidissima*	A	71	10.5
223	5	BLP	*S.s. solidissima*	A	71	3
224	5	BLP	*S.s. solidissima*	A	71	3.4
225	5	BLP	*S.s. solidissima*	A	71	3.2
226	5	BLP	*S.s. solidissima*	A	72	3.5
227	5	BLP	*S.s. solidissima*	A	72	3.5
228	5	BLP	*S.s. solidissima*	unsequenced	73	
229	5	BLP	*S.s. solidissima*	unsequenced	73	
230	5	BLP	*S.s. solidissima*	unsequenced	73	
231	5	BLP	*S.s. solidissima*	A	73	7.4
232	5	BLP	*S.s. solidissima*	A	73	3.4
233	5	BLP	*S.s. solidissima*	A	74	6.3
234	5	BLP	*S.s. solidissima*	A	74	7.8
235	5	BLP	*S.s. solidissima*	A	75	11.5
236	5	BLP	*S.s. solidissima*	unsequenced	75	
237	5	BLP	*S.s. solidissima*	A	75	7.5
238	5	BLP	*S.s. solidissima*	unsequenced	75	
239	5	BLP	*S.s. solidissima*	A	75	3.4

240	5	BLP	*S.s. solidissima*	A	75	3
241	5	BLP	*S.s. solidissima*	A	75	2.7
242	5	BLP	*S.s. solidissima*	A	75	3
243	5	BLP	*S.s. solidissima*	A	75	1.6
244	6	CUP	*S.s. solidissima*	unsequenced	76	7.3
245	6	CUP	*S.s. solidissima*	unsequenced	76	6.7
246	6	CUP	*S.s. solidissima*	unsequenced	76	12.5
247	6	CUP	*S.s. solidissima*	unsequenced	77	7
248	6	CUP	*S.s. solidissima*	unsequenced	77	6.9
249	6	CUP	*S.s. solidissima*	unsequenced	77	7
250	6	CUP	*S.s. solidissima*	unsequenced	77	6.5
251	6	CUP	*S.s. solidissima*	unsequenced	77	7.1
252	6	CUP	*S.s. solidissima*	unsequenced	77	
253	6	CUP	*S.s. solidissima*	unsequenced	77	
254	6	CUP	*S.s. solidissima*	unsequenced	78	6.9
255	6	CUP	*S.s. solidissima*	unsequenced	78	7.3
256	6	CUP	*S.s. solidissima*	unsequenced	78	6.9
257	6	CUP	*S.s. solidissima*	unsequenced	78	6.6
258	6	CUP	*S.s. solidissima*	unsequenced	78	6.6
259	6	CUP	*S.s. solidissima*	unsequenced	78	
260	6	CUP	*S.s. solidissima*	unsequenced	78	
261	6	CUP	*S.s. solidissima*	unsequenced	78	
262	6	CUP	*S.s. solidissima*	unsequenced	78	
263	6	CUP	*S.s. solidissima*	unsequenced	78	
264	6	CUP	*S.s. solidissima*	unsequenced	78	
265	6	CUP	*S.s. solidissima*	unsequenced	78	
266	6	CUP	*S.s. solidissima*	unsequenced	78	
267	6	CUP	*S.s. solidissima*	unsequenced	78	
268	6	CUP	*S.s. solidissima*	unsequenced	78	
269	6	CUP	*S.s. solidissima*	unsequenced	79	
270	6	CUP	*S.s. solidissima*	unsequenced	79	
271	6	CUP	*S.s. solidissima*	unsequenced	79	
272	6	CUP	*S.s. solidissima*	unsequenced	79	
273	6	CUP	*S.s. solidissima*	unsequenced	79	
274	6	CUP	*S.s. solidissima*	unsequenced	79	

275	6	CUP	*S.s. solidissima*	unsequenced	79	
276	6	CUP	*S.s. solidissima*	unsequenced	79	12
277	6	CUP	*S.s. solidissima*	unsequenced	79	7.2
278	6	CUP	*S.s. solidissima*	unsequenced	79	7.4
279	6	CUP	*S.s. solidissima*	unsequenced	79	6.4
280	6	CUP	*S.s. solidissima*	unsequenced	79	7.1
281	6	CUP	*S.s. solidissima*	unsequenced	79	6.5
282	6	CUP	*S.s. solidissima*	unsequenced	80	
283	6	CUP	*S.s. solidissima*	unsequenced	80	8
284	6	CUP	*S.s. solidissima*	unsequenced	80	9.2
285	6	CUP	*S.s. solidissima*	unsequenced	81	8
286	6	CUP	*S.s. solidissima*	unsequenced	81	
287	6	CUP	*S.s. solidissima*	unsequenced	81	
288	6	CUP	*S.s. solidissima*	unsequenced	81	
289	6	CUP	*S.s. solidissima*	unsequenced	81	
290	6	CUP	*S.s. solidissima*	unsequenced	81	
291	6	CUP	*S.s. solidissima*	unsequenced	81	
292	6	CUP	*S.s. solidissima*	unsequenced	81	
293	6	CUP	*S.s. solidissima*	unsequenced	82	8.5
294	6	CUP	*S.s. solidissima*	unsequenced	82	6
295	6	CUP	*S.s. solidissima*	unsequenced	82	
296	6	CUP	*S.s. solidissima*	unsequenced	82	
297	6	CUP	*S.s. solidissima*	unsequenced	82	
298	6	CUP	*S.s. solidissima*	unsequenced	82	
299	6	CUP	*S.s. solidissima*	unsequenced	82	
300	6	CUP	*S.s. solidissima*	unsequenced	82	
301	6	CUP	*S.s. solidissima*	unsequenced	82	
302	6	CUP	*S.s. solidissima*	unsequenced	82	
303	6	CUP	*S.s. solidissima*	unsequenced	82	
304	6	CUP	*S.s. solidissima*	unsequenced	83	5.5
305	6	CUP	*S.s. solidissima*	unsequenced	83	8.1
306	6	CUP	*S.s. solidissima*	unsequenced	83	7.3
307	6	CUP	*S.s. solidissima*	unsequenced	83	
308	6	CUP	*S.s. solidissima*	unsequenced	83	2.9
309	6	CUP	*S.s. solidissima*	unsequenced	83	2.6

310	6	CUP	*S.s. solidissima*	unsequenced	83	3.1
311	6	CUP	*S.s. solidissima*	unsequenced	83	3.4
312	6	CUP	*S.s. solidissima*	unsequenced	83	3.3
313	6	CUP	*S.s. solidissima*	unsequenced	83	2.4
314	6	CUP	*S.s. solidissima*	unsequenced	83	2.5
315	6	CUP	*S.s. solidissima*	unsequenced	84	6.8
316	6	CUP	*S.s. solidissima*	unsequenced	84	6.5
317	6	CUP	*S.s. solidissima*	unsequenced	84	8.5
318	6	CUP	*S.s. solidissima*	unsequenced	84	6.7
319	6	CUP	*S.s. solidissima*	unsequenced	84	7
320	6	CUP	*S.s. solidissima*	unsequenced	84	7.7
321	6	CUP	*S.s. solidissima*	unsequenced	84	7.8
322	6	CUP	*S.s. solidissima*	unsequenced	84	7.1
323	6	CUP	*S.s. solidissima*	unsequenced	84	6.6
324	6	CUP	*S.s. solidissima*	unsequenced	84	7.4
325	6	CUP	*S.s. solidissima*	unsequenced	84	8.5
326	6	CUP	*S.s. solidissima*	unsequenced	84	
327	6	CUP	*S.s. solidissima*	unsequenced	84	
328	6	CUP	*S.s. solidissima*	unsequenced	84	
329	6	CUP	*S.s. solidissima*	unsequenced	84	
330	6	CUP	*S.s. solidissima*	unsequenced	84	
331	6	CUP	*S.s. solidissima*	unsequenced	84	
332	6	CUP	*S.s. solidissima*	unsequenced	85	6.5
333	6	CUP	*S.s. solidissima*	unsequenced	85	6.7
334	6	CUP	*S.s. solidissima*	unsequenced	85	7.2
335	6	CUP	*S.s. solidissima*	unsequenced	85	5.1
336	6	CUP	*S.s. solidissima*	unsequenced	85	3.5
337	6	CUP	*S.s. solidissima*	unsequenced	85	3.2
338	6	CUP	*S.s. solidissima*	unsequenced	85	6.6
339	6	CUP	*S.s. solidissima*	unsequenced	85	
340	6	CUP	*S.s. solidissima*	unsequenced	85	
341	6	CUP	*S.s. solidissima*	unsequenced	85	
342	6	CUP	*S.s. solidissima*	unsequenced	85	
343	6	CUP	*S.s. solidissima*	unsequenced	85	
344	6	CUP	*S.s. solidissima*	unsequenced	85	

345	6	CUP	*S.s. solidissima*	unsequenced	85	
346	6	CUP	*S.s. solidissima*	unsequenced	85	
347	6	CUP	*S.s. solidissima*	unsequenced	86	6.6
348	6	CUP	*S.s. solidissima*	unsequenced	86	7.2
349	6	CUP	*S.s. solidissima*	unsequenced	86	8.3
350	6	CUP	*S.s. solidissima*	unsequenced	86	6.2
351	6	CUP	*S.s. solidissima*	unsequenced	86	6.6
352	6	CUP	*S.s. solidissima*	unsequenced	86	2.5
353	6	CUP	*S.s. solidissima*	unsequenced	86	2.6
354	6	CUP	*S.s. solidissima*	unsequenced	86	2.8
355	6	CUP	*S.s. solidissima*	unsequenced	86	2.4
356	6	CUP	*S.s. solidissima*	unsequenced	86	2.5
357	6	CUP	*S.s. solidissima*	unsequenced	86	2.1
358	6	CUP	*S.s. solidissima*	unsequenced	86	2.6
359	6	CUP	*S.s. solidissima*	unsequenced	86	3.1
360	6	CUP	*S.s. solidissima*	unsequenced	86	
361	6	CUP	*S.s. solidissima*	unsequenced	86	
362	6	CUP	*S.s. solidissima*	unsequenced	86	
363	6	CUP	*S.s. solidissima*	unsequenced	86	
364	6	CUP	*S.s. solidissima*	unsequenced	86	
365	6	CUP	*S.s. solidissima*	unsequenced	86	
366	6	CUP	*S.s. solidissima*	unsequenced	86	
367	6	CUP	*S.s. solidissima*	unsequenced	86	
368	6	CUP	*S.s. solidissima*	unsequenced	86	
369	6	CUP	*S.s. solidissima*	unsequenced	86	
370	6	CUP	*S.s. solidissima*	unsequenced	86	
371	6	CUP	*S.s. solidissima*	unsequenced	86	5.6
372	6	CUP	*S.s. solidissima*	unsequenced	86	
373	6	CUP	*S.s. solidissima*	unsequenced	86	
374	6	CUP	*S.s. solidissima*	unsequenced	86	
375	6	CUP	*S.s. solidissima*	unsequenced	86	
376	6	CUP	*S.s. solidissima*	unsequenced	86	
377	6	CUP	*S.s. solidissima*	unsequenced	87	6.1
378	6	CUP	*S.s. solidissima*	unsequenced	87	5
379	6	CUP	*S.s. solidissima*	unsequenced	87	1.9

380	6	CUP	*S.s. solidissima*	unsequenced	87	12.9
381	6	CUP	*S.s. solidissima*	unsequenced	87	12.7
382	6	CUP	*S.s. solidissima*	unsequenced	87	
383	6	CUP	*S.s. solidissima*	unsequenced	87	
384	6	CUP	*S.s. solidissima*	unsequenced	87	
385	6	CUP	*S.s. solidissima*	unsequenced	87	
386	6	CUP	*S.s. solidissima*	unsequenced	87	
387	6	CUP	*S.s. solidissima*	unsequenced	87	
388	6	CUP	*S.s. solidissima*	unsequenced	87	
389	6	CUP	*S.s. solidissima*	unsequenced	87	
390	6	CUP	*S.s. solidissima*	unsequenced	87	
391	6	CUP	*S.s. solidissima*	unsequenced	87	
392	6	CUP	*S.s. solidissima*	unsequenced	87	
393	6	CUP	*S.s. solidissima*	unsequenced	88	
394	6	CUP	*S.s. solidissima*	unsequenced	88	3.5
395	6	CUP	*S.s. solidissima*	unsequenced	88	
396	6	CUP	*S.s. solidissima*	unsequenced	88	
397	6	CUP	*S.s. solidissima*	unsequenced	88	
398	6	CUP	*S.s. solidissima*	unsequenced	88	
399	6	CUP	*S.s. solidissima*	unsequenced	89	3
400	6	CUP	*S.s. solidissima*	unsequenced	89	
401	6	CUP	*S.s. solidissima*	unsequenced	90	
402	6	CUP	*S.s. solidissima*	unsequenced	90	
403	6	CUP	*S.s. solidissima*	unsequenced	90	
404	6	CUP	*S.s. solidissima*	unsequenced	90	
405	6	CUP	*S.s. solidissima*	unsequenced	90	3