AN EXAMINATION OF *CRASSOSTREA VIRGINICA* NUCLEAR DNA VARIATION ALONG THE NORTH CAROLINA COAST

Russ Peterson

A Thesis Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
of the Requirements for the Degree of
Master of Science

Department of Biology and Marine Biology

University of North Carolina Wilmington

2006

Approved by

Advisory Committee

Dr. Michael McCartney

Dr. Martin Posey

Dr. Ami Wilbur
Chair

Accepted by

_____
Dean, Graduate School

TABLE OF CONTENTS

ABSTRACT

Inhabiting coastal waters from eastern Canada to the Gulf of Mexico, the eastern oyster is subjected to a wide range of temperature and salinity regimes, thus providing an interesting opportunity to study population structure. Prior studies have examined phenotypic as well as DNA differences along this range. A previous mtDNA population survey of *Crassostrea virginica* within Pamlico Sound utilizing a single 16s polymorphism diagnostic for North Atlantic / South Atlantic haplotypes revealed an ~110 km cline along the North Carolina coast. Using 4 microsatellite loci, 3 SNPs and 1 scnDNA RFLP, I have surveyed eight oyster populations within and outside the Pamlico Sound in an effort to corroborate the population structure found in the mitochondrial genome. Three microsatellite loci were out of HWE across populations vs. only 1 population for one SNP loci, and it seems likely that those microsatellite loci were plagued with null alleles. Microsatellite exact tests show some significant differences within the Pamlico Sound, mostly in comparisons involving the Stumpy Point population. A combined SNP/RFLP analysis did reveal significant differences among populations, though most of this can be accounted for by inclusion of a population from Maryland. The clinal structure seen in the mitochondrial genome is not reflected in the nuclear genome within the Pamlico Sound.

ACKNOWLEDGEMENTS

I am indebted to my many lab mates over the years, including Greg Hoffman, Rachel Sackett, Alison Slocum, Bob York, Sam Honeycutt, and Elizabeth Hemond.

Also, I like to thank many others outside the lab for their general contribution to my life, including Christin Slaughter, Kristi Sommer, Chip Collier, Monica McGee, Craig Bailey, Ann Stapleton, Ann Pabst, and Wilson Freshwater.

Thanks also to my committee members, Mike McCartney and Martin Posey, as well as Ami Wilbur, for her never ending supply of patience.

List of Tables

List of Figures

Figure Page

INTRODUCTION

Crassostrea virginica is a euryhaline, eurythermal species that inhabits the North American coast along both the Gulf of Mexico and the Atlantic Ocean (Buroker, 1983). It has a planktonic larval stage that can last up to three weeks and thus is thought to be capable of long-distance dispersal. As a consequence, high gene flow is expected among geographic populations resulting in low levels of population structure (Hedgecock, 1982). Many studies, however, have suggested that oysters exhibit unexpected differences among geographically separated populations.

Loosanoff and Nomejko (1951) detected differences in spawning cues among oysters from five different geographical locations transplanted in Milford, Connecticut during the summer of 1948. Oysters from Massachusetts completely discharged all gametes during the summer of 1950. Lack of spawning or incomplete spawning was seen in oysters originating in the warmer waters of New Jersey and Virginia. A high rate of over-winter mortality of North Carolina and Florida oysters was also noted (87-97%). They concluded that there may be oyster races adapted to particular thermal regimes that differ throughout their range.

A transplant study by Barber et al. (1991) provided additional evidence of geographic differentiation. They examined the spawning and maturation of gonadal tissue in inbred transplanted and native strains of oysters from Long Island Sound and Delaware Bay that had been propagated in Delaware Bay. Their aim was to eliminate physiological acclimatization as a factor in the spawning and gonadal maturation cycle, thereby determining the contribution of genetic versus non-genetic factors previously observed in the phenotypic differences seen in spawning. They found that the geographic differences in timing of spawning in relation to water temperature were maintained in the Long Island Sound group after 23 years (6 generations) in

the Delaware Bay. Long Island Sound oysters began spawning at water temperatures between 15-20 ºC, whereas 25-28 ºC stimulated spawning in the Delaware Bay oysters. However, these authors also note that oysters transplanted from Maryland to Florida (Butler, 1955, as cited by Barber et al., 1991) conformed to local spawning patterns after one year.

Another transplant study examined the effects of year class, origin, and age on growth patterns using oysters from Long Island Sound, Delaware Bay, and James River (Dittman, et al. 1998). The authors point out that there exists a positive correlation between body size and the circumvention of predators. A correlation also exists with reproductive output as well as with environmental factors, such as average water temperature and the supply of food, as major determinates of growth rate. They found that Long Island Sound oysters raised in the common environment with the Delaware Bay and James River samples were consistently larger than oysters from both of the other populations, which could signify the existence of stocks under localized selection. They suggest that temperature dependent physiological functions, such as feeding mechanisms (timing, duration and digestive efficiency of feeding) as well as the apportioning of energy resources (growth rate vs. reproduction), may be rooted in strain origin.

Hillman (1964) used paper partition chromatography to evaluate differences in free amino acid pools in oysters from New York (Long Island Sound) and Virginia (James River). He found the pool composition to vary consistently among populations, and attributed the differences to the specific salinity regimes experienced by the oysters in their natal sites. He suggested that salinity as well as temperatures may be important in the development of geographic races of oysters.

In contrast to this evidence suggestive of physiological differentiation, an extensive survey of allozyme variation revealed little differentiation among geographic populations.

Buroker (1983) sampled oysters from 19 locations along the Atlantic coast and the Gulf of Mexico and subjected them to protein electrophoresis using 23 protein-staining systems. This study found that the only population that exhibited significant differentiation was Brownsville, Texas, which was the southern-most site. Buroker thus concluded that oysters were genetically homogenous throughout their range. A re-evaluation of these data was performed by Cunningham and Collins (1994) using phylogenetic approaches. Using a variety of tree building methodologies, oysters from Atlantic and peninsular Florida localities consistently formed monophyletic groups distinct from oysters from the Gulf of Mexico (with the break occurring off of western Florida) which is indicative of significant population structure between these regions.

These seemingly contradictory observations of physiological variation and extensive protein homogeneity among geographic populations of oysters have led to a number of additional studies using increasingly more sensitive techniques and analysis.

Analysis of mitochondrial DNA (mtDNA) variation detected significant differentiation of Atlantic and Gulf populations (Reeb and Avise, 1990) as was revealed by the Cunningham and Collins (1994) reanalysis of Buroker's (1983) allozyme work. This study evaluated restriction fragment length polymorphisms (RFLP) in mtDNA of oysters collected from the coast of Canada to the coast off Texas. They found that restriction patterns differed significantly between Gulf and Atlantic oysters, and their analysis suggests that the oyster falls into at least two distinct assemblages. While this supports the reanalysis of Buroker's (1983) work by Cunningham and Collins (1994) with regard to Gulf and Atlantic assemblages, the two studies vary in the location of the break, and this does not provide any insight into the physiological differences previously reported within the Atlantic basin, as neither the allozyme nor the mtDNA data analyses suggest any within Atlantic differentiation.

The differentiation of Gulf and Atlantic populations has been reported along eastern Florida for a variety of species (Avise, 1992), which led Reeb and Avise (1990) to the conclusion that a vicariant event occurring in the Pleistocene period is responsible for the divergence in oyster mtDNA (Rezak et al., 1985). Historical vicariant events, when contiguous ranges are separated by some geophysical or ecological event, are used as one hypothesis to explain current morphological or genetic disjuncts in related taxa (Avise, 2004). Similar patterns of differentiation in mtDNA variation have been observed in 19 species of freshwater, coastal, and marine species in the southeastern U.S. (Avise, 1992). Such strong concordance of data across such a wide variety of species is considered to be evidence for a vicariant break. Karl and Avise (1992) followed up Reeb and Avise's (1990) work by trying to evaluate the extent of concordance between the mitochondrial and the nuclear genomes by evaluating allele frequency variation at nuclear markers (RFLPs at single copy nuclear DNA (scnDNA)) in oysters from sampling sites stretching from Massachusetts to Louisiana. Their data showed concordance with Reeb and Avise's (1990) earlier study by confirming the two assemblages and placing the genetic break off the coast of eastern Florida at Cape Canaveral.

Karl and Avise's (1992) results were questioned by McDonald et al. (1996), who analyzed variation at six anonymous nuclear DNA (nDNA) loci in samples taken from Charleston, South Carolina (Atlantic) and Panacea, Florida (Gulf) and found no significant differences. They suggest that the inconsistencies in the data might be explained by the possibility that Karl and Avise's (1992) technique (scnDNA) might preferentially identify areas of DNA that are geographically differentiated, as well as by the presence of null (non-amplifying) alleles. In 1996, Hare and Avise revisited the question, surveying oysters from 18 different sites using RFLPs of mtDNA and two of Karl and Avise's (1992) scnDNA loci (Karl

and Avise, 1992), using internal primers. Their data, again supported the existence of a genetic break near Cape Canaveral, Florida. Although no plausible explanation was given for the discrepancies between their results and McDonald's et al. (1996) results, Hare and Avise (1996) present possible explanations for maintenance of the break, including marine currents, selection processes, and water mass differences.

This presence of a genetic break was further supported by sequence data from three nuclear loci (Hare and Avise, 1998). Adult oysters were collected from two locations in the Atlantic (Massachusetts and South Carolina) as well as from two locations in the Gulf (western Florida and Louisiana). Although some alleles were found only within either the Atlantic or Gulf locales, reciprocal monophyly of the data set as a whole was not observed for any of the loci tested. They suggest that mtDNA, due to its haploid nature as well as the maternal mode of transmission, may have had time to sort into distinctive lineages, whereas nDNA had not.

A second genetic break in oysters was uncovered by a study that used denaturing gradient gel electrophoresis (DGGE) to reanalyze the oysters used in the Reeb and Avise (1990) study (Wakefield, 1996). Two nucleotide substitutions in the 16s ribosomal subunit effectively separated oyster populations into Gulf, South Atlantic and North Atlantic stocks. The Gulf haplotype dominated south of the Cape Canaveral break previously observed, but a second significant shift in frequencies of North and South Atlantic haplotypes was also detected at Cape Hatteras, NC. Sackett (2002) investigated the purported North Carolina break by sequencing ~ 400 base pairs (bp) of the 16s ribosomal subunit for several sites along the North Carolina coast and revealed a pattern of decreasing frequency of South Atlantic haplotypes and increasing frequency of North Atlantic haplotypes along 110 kilometers of the NC coast (Figure 1). This

pattern was further substantiated using RFLP analysis on the COIII and ATPase 6 mtDNA loci, which also differentiate North Atlantic and South Atlantic oysters (Gaffney, pers com).
As was the case for the genetic break off of Florida, the discontinuity off of North Carolina is unexpected, given the continuous distribution of oysters and their dispersive, planktonic larval form. Possible explanations for this North Carolina genetic break are similar to those presented for the Cape Canaveral break. First, a vicariant event that isolated North and South assemblages could have allowed the separated groups to diverge. Secondary contact following a breakdown of the isolation event and limits on effective larval dispersal could explain the clinal pattern observed in Sackett (2002). Neutral nuclear DNA markers, however, may or may not show the same pattern depending on the duration of the isolation. Nuclear DNA, being both diploid and having a greater effective population size, may not have had sufficient time either to drift to fixation or to drift sufficiently to show differentiated frequencies in separate assemblages.

Alternatively, genetic differentiation could be the result of selection operating in conjunction with drift acting on nuclear loci. For example, temperature or salinity gradients affecting the timing of spawning could emerge as a form of pre-zygotic isolation producing locally adapted regional stocks, which would allow mtDNA, and possibly neutral nDNA, to quickly sort into distinct lineages. Further work needs to be done in this area utilizing both the mtDNA haplotypes as well as including nDNA. The use of nDNA could provide a clearer picture of the distribution of stocks as it would include the paternal genetic contribution. The identification of population structure should help illuminate past and present pressures. For instance, a geographic distribution of nDNA haplotypes similar to the mtDNA haplotypes would be suggestive of admixture of assemblages, rather than introgression of alleles due to secondary

contact. The inclusion of nDNA should also help assess the amount of genetic variability along the North Carolina coast.

Microsatellites, genetic markers used for their supposed neutrality and high levels of heterozygosity (Hedrick 1999), have been successfully developed for oysters (Reece et al. 2004, Brown et al. 2000) and have been used to examine gene flow in the Chesapeake Bay (Rose et al. 2006,).

Single nucleotide polymorphisms (SNPs) in both the nuclear and mitochondrial genomes have been used extensively in the past to glean information from organisms about gene flow and population structure in oysters (Reeb and Avise 1990, Karl and Avise, 1992, Sackett 2002). The technique known as single base extension (SBE), in which a gene with an identified SNP is amplified through the polymerase chain reaction and then targeted with a single primer and subsequent reamplification (Applied Biosystems), will be used in this study.

In this study I evaluate the extent of genetic differentiation near or along the coast of North Carolina by examining nuclear variation in oysters utilizing four microsatellite loci, 3 SNPs from three different loci developed from an EST database, as well as one scnDNA locus subjected to RFLP. Data are analyzed both according to previous mtDNA assignment (Sackett 2002), for concordance between genomes, as well as by geographic location.

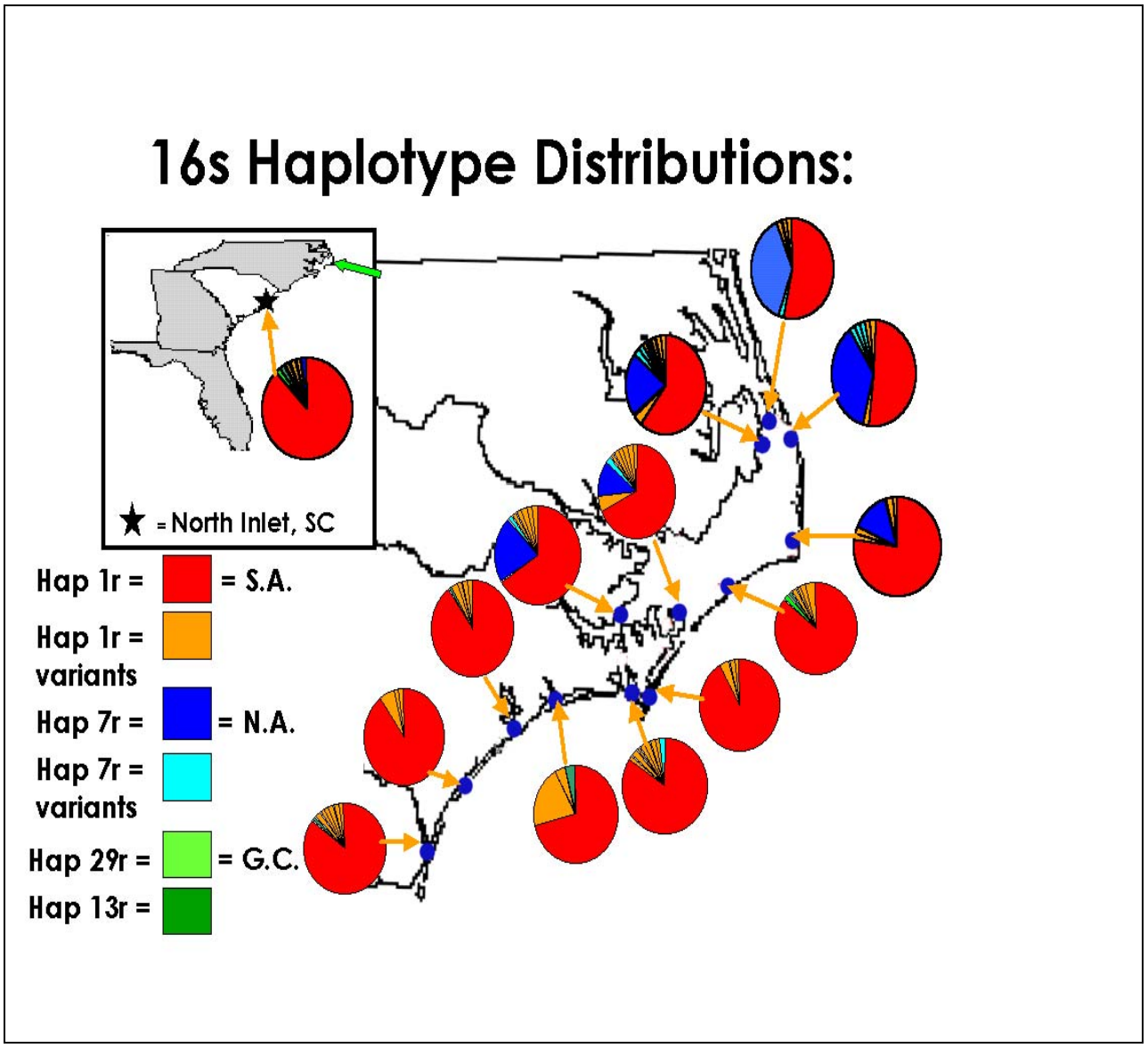

Figure 1. 16s haplotype distribution in oysters based on sequencing of 450bp. S.A. = South Atlantic, N.A. = North Atlantic and G.C. = Gulf Coast (Sackett 2002). Regional haplotypes are described in Wakefield (1996).

MATERIALS AND METHODS

Oyster Samples and DNA Extraction

I used oyster samples collected in September 2001 from 6 locations in the Pamlico Sound, one from South Carolina, and one from Maryland (Figure 2). These same samples were subject to mtDNA analysis (with the exception of the Maryland sample) and the results reported in Sackett (2002). The original DNA extractions, as well as any needed replicate extractions were derived from adductor muscle using the Purgene™ extraction kit according to manufacturer's protocols scaled down for use with small amounts of tissue (Gentra Systems, Minneapolis MN). Extracts were then stored at -20°C and diluted 1:8 to create working stocks for amplification. Dilutions were stored at 4° C.

Microsatellites

Four microsatellite loci (Table 1) were amplified on either an MJ Research 100 or 200 thermalcycler using pig-tailed primers to reduce stuttering (Brownstein et al., 1996). Cvi2i23 and Cvi2j24 are tetranucleotide repeats chosen due to their conformation to expectations given Mendelian inheritance during single pair crosses as determined by analysis of offspring from Reece et al. (2004). Cvi6 and Cvi13 are trinucleotide and tetranucleotide repeats, respectively, that amplified well in previous studies (Brown et al., 2000). Products were visualized on a 2% agarose/ETBR gel to verify amplification. Amplicons were then diluted 1:100 with deionized water (dH_2O), with 1 µl of the dilution added to 9 µl's of a Genescan™ -500 ROX™ size standard/Hi-Di™ Formamide mix, and electrophoresed on an ABI3100 Genetic Analyzer. Results were evaluated and sized using Genotyper™ Version 3.7.

Single Nucleotide Polymorphisms (SNPs)

Amplification primers (Table 2) for the chitinase, RAN, and glutathione peroxidase genes were designed from an EST database (CvCHI and CvRAN, Varney and Gaffney, pers. comm.; CvGP, Wilbur, pers. comm.) Each locus was sequenced and flanking regions were examined for variation using a subset of oysters from sites along the Atlantic Coast. Candidate SNPs were identified, and then primers designed for use in Single Base Extension (SBE) assays. The primer CH162 was used to detect a A/C transversion at base position 162 while the RANB SNP primer was developed to detect an A/G transition at base position 100 (Gaffney, pers. com). The GP175 SNP primer was designed to target a G/C transversion observed at position 175.

All target products were amplified using a Promega *Taq* DNA Polymerase kit on either an MJ100 or MJ200 PCR. All reactions included 1 μl diluted DNA, 0.15 μl *Taq* polymerase, 1 μl each primer (10mM), and 2.5 ml 10x buffer. CvRAN and CvGP contained 2.5 μl each of $MgCl_2$ and dNTPs (2 mM), while CvCHI contained 2.0 μl each of $MgCl_2$ and dNTPs (2mM), with adequate dH_2O added to bring reaction volume to 25 μl. All PCR products were run on 2% agarose gels stained with ETBR to verify amplification. SNPs were then targeted using SNP primers (Table 1.b) and an ABI Prism^{R} SNaPshot^{TM} Multiplex Kit using a modified protocol. PCR products (5μl) were mixed with 2 μl of EXOSAP-IT^{R} (USB Corp. Cleveland, Ohio) and incubated at 37^{o} C for 1 hour, followed by 30 minutes at 80^{o} C. Then, SNP reactions were set up using 0.625 μl SNaPshot^{TM} Multiplex reaction mix, 0.625 μl 2.5X dilution buffer, 0.25 μl SNP primer, 0.25 μl dH_2O and 0.75 μl of cleaned PCR. Reactions were then amplified under the

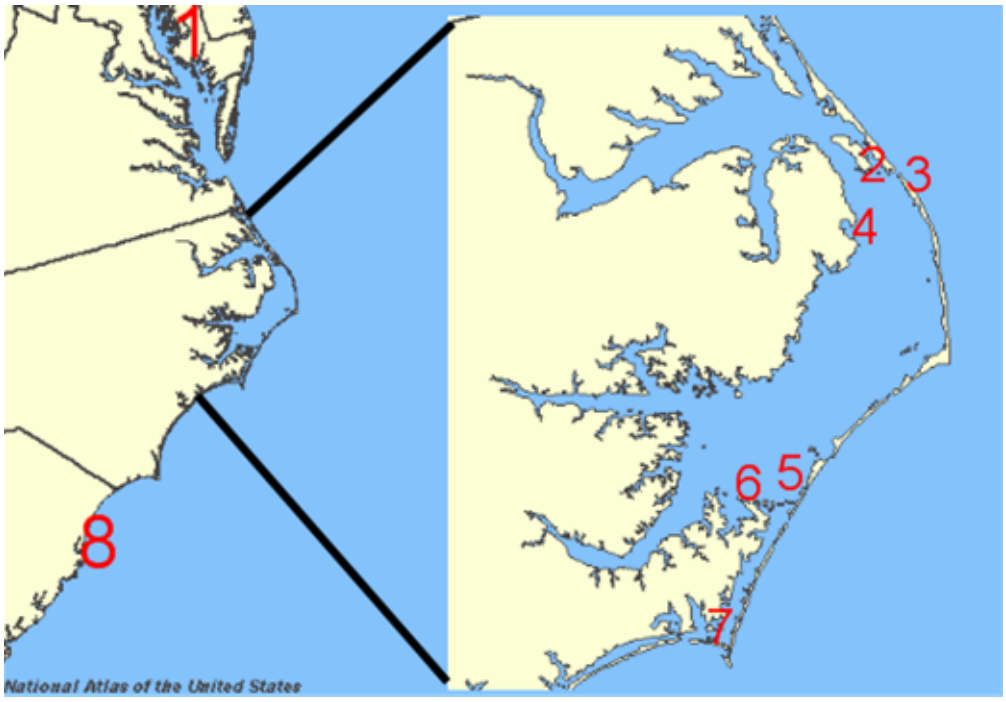

Figure 2. Locations and abbreviations for sites used in this study. 1) Choptank River, MD, (CH), 2) Wanchese, NC, (WA), 3)Oregon Inlet, NC, (OI), 4) Stumpy Point, NC, (SP), 5) Ocracoke Island, NC (OC), 6) Cedar Island, NC (CI), 7) Harker's Island, NC, (HI), 8) North Inlet, SC, (SC).

following conditions: 25 cycles of 96^{o} C for 10 s., 50^{o} C for 5s., and 60^{o} C for 30 s., followed by a 4^{o} C hold. A post-amplification clean up was done by adding 0.5 μl Shrimp Alkaline Phosphatase (SAP; USB Corp. Cleveland, Ohio) diluted 1:2 with SAP dilution buffer to each 2.5 μl sample and incubating at 37^{o} C for 90 minutes. Then 0.5 μl of cleaned reaction was then added to 9 μl Hi-Di^{TM} Formamide and 0.5 μl GeneScan^{TM} LIZ^{R} size standard, denatured at 95^{o} C for 2 minutes, placed on ice for ~ 5 minutes, then run on an ABI3100 Genetic Analyzer and viewed using Genescan^{TM} Version 3.7.

Restriction Fragment Lengths Polymorphisms (RFLP)

Cv195 (Table 1), a single copy nuclear locus developed by Karl and Avise (1992), was amplified and sequenced in a subset of animals to identify potential restriction sites. Six individuals (01WN-01, -12, -15, -18, -19, and -25) were cloned (60 clones) using a Promega pGEM^{TM}-T Easy Vector System kit according to manufacturer's protocol. Clones were then sequenced and examined in Sequencher^{TM} for polymorphisms. A transition (G/A) was observed at position 177, which resulted in an RFLP site detectable using the enzyme *Dde1*. An internal primer was designed to anneal in a conserved area for the consistent amplification of a smaller pcr product (~242bp) containing the restriction site. *Dde1* restriction enzyme (NEB Beverly, Massachusetts) cuts either once or twice on the resultant Cv195 amplicon. All amplicons cut at a common position which results in an ~60bp restriction fragment. The variable site cleaves the residual fragment (177bp) into two fragments (133bp and 44bp) in amplicons with a G at position 177. All digests were run according to NEB protocol, visualized with ETBR stained 3% gels, and scored by eye.

Statistical Analysis

All data were recorded in Excel and converted to appropriate statistical package formats using Convert (Glaubitz, 2004). Microsatellites were checked for large allele drop out, stuttering, and null alleles using MICRO-CHECKER Version 2.2.3 (Van Oosterhout et al., 2004). Hardy Weinberg Equilibrium (HWE, exact test, Markov Chain method), linkage disequilibrium, population differentiation (genic differentiation, Fisher exact test) and Isolation by Distance (IBD, Isolde) were tested using GENEPOP (Raymond and Rousset, 1995). Arlequin version 3.01 (Excoffier et. al., 2000) was used to examine population structure through an AMOVA (Excoffier et al., 1992) as well as to compute pairwise F_{ST} (Weir and Hill, 2002). Sequential Bonferroni (Rice, 1989) correction has been applied to P-values across tables where appropriate. Pertinent analyses were configured by 16s haplotype as well as by location.

Table 1. Microsatellite amplicon length, annealing temperature, allele count and primer source.

Microsatellite Loci	Product Length (base pairs)	Annealing T°C	Allele Number	Primer Sequences	
Cvi2i23	362-606	53°	40	Reece et. al. (2004)	
Cvi2j24	367-439	53°	16	Reece et. al. (2004)	Cvi6
145-235	50°	25	Brown et. al. (2000)		
Cvi13	117-309	50°	37	Brown et. al. (2000)	

Table 2. SNP loci, annealing temperature, external primers and internal SNP primer.

SNP Loci	Annealing T°C	Initial Primer Sequences and SNP Primer
CvGP	50°	CvGP-F 5'-GCA AGC CAT TCC GAA GAT AC-3'
		CvGP-R 5'-TTG CGT GCA CAT CAT ATC CT-3'
		GP175 5'-GTG GAA GAA TAA TCA TTA AAG CAA AT-3'
CvRAN	50°	CvRAN-F 5'-AAA TGT TCC CAA CTG GCA TAG AGA-3'
		CvRAN-R 5'-CTC CCA CCA ATT TCC TAG CTA ACC-3'
		RANB 5'-CGA CAT CAA GGA TCG CAA AGT TAA-3'
CvCHI	63°	CvCHI-F 5'-CGG CAG AGT ACT GGC ACC AGA AGG-3'
		CvCHI-R 5'-CGT TAT TGC TCC CGG AAA TG-3'
		CH162 5'-GCT AGC CCG GCC CCT GAC AC-3'

Table 3. RFLP locus, annealing temperature, original primer sequence and revised internal primer sequence.

RFLP Locus	Annealing T°C	Primer Sequences
Cv195	65°	Cv195-Fint 5'-TAC AAA TTC TTT TCG CATTCC CAW G-3'
		Cv195-F 5'-GGA TCA GAA GGA AAG CAA CAG CAC-3'
		Cv195-R 5'-AAC GTT TGA TGG AAC AAG GGA AAC T-3'

RESULTS

Microsatellites

Cvi2i23 amplified across a large range of allele sizes (362-606 base pairs) with 40 separate alleles found in all populations. There was evidence for a large indel within the locus, as is suggested by the bimodal distribution seen in the data, but it did not appear to be restricted to any particular population. Cvi2j24, Cvi6 and Cvi13 all had a smaller size range and fewer alleles. No significant linkage disequilibrium was detected ($P>0.05$).

No significant deviation from Hardy-Weinberg equilibrium was seen for Cvi2i23. Analysis of the Cvi2j24, Cvi6 and Cvi13 loci revealed significant departures from equilibrium for multiple populations (Table 4), all in the direction of heterozygote deficiencies. Evaluation of genotype frequencies suggests the presence of null alleles within these loci, and genotype frequencies were adjusted accordingly (MICROCHECKER, Van Oosterhout, 2004). Re-analysis of Hardy-Weinberg equilibrium for these loci revealed fewer deviations (Table 4).

Population genic and genotypic differences were computed using an exact test as implemented in GENEPOP for each locus, as well as a global test, analyzed by geography and 16s haplotype (Tables 5-10). There were more significant differences between locations found at the genic level for the loci than at the genotypic level, and most of the differences seemed to be within the Pamlico Sound. None of the microsatellite exact tests analyzed by 16s were significant.

IBD was tested in GENEPOP using the Isolde program and the original data. None of the tests of regressions plotted against distance for any of the four loci were significant (Figure 3).

An AMOVA was performed using all four loci partitioned into both geographical populations as well as by individual oyster 16s haplotype (Table 11). Among populations percentage of variation accounted for by both partitions ranged from 0.05 to 0.17%, neither of which was significant. Pairwise differences were also examined by location (Table 12) and 16s haplotype (Φ_{ST}: 0.0009, P-value: 0.2646). Again, no significant differences were found.

Table 4. Microsatellite loci HWE P-values by location. H_e and H_o are percentage of heterozygous individuals expected and observed, respectively; P is the P-value from HWE; (cor) is the P-value from HWE using adjusted genotype frequencies from MICRO-CHECKER where appropriate (- indicates no evidence for null alleles); N is the number of individuals amplified with the respective loci. Significant P-values after sequential Bonferroni correction in bold.

	Cvi2i23	Cvi2j24	Cvi6	Cvi15
Choptank River				
H_e	0.9200	0.8526	0.8390	0.9070
H_o	0.9130	0.5200	0.6040	0.8160
P	0.9177	**0.0000**	**0.0001**	**0.0134**
S.E.	0.0040	0.0000	0.0001	0.0018
cor	--------	**0.0057**	**0.0069**	--------
S.E.	--------	0.0003	0.0010	--------
N	50	50	48	49
Wanchese				
H_e	0.9150	0.8760	0.8700	0.9090
H_o	0.9330	0.6520	0.5600	0.7920
P	0.7936	**0.0093**	**0.0000**	0.2456
S.E.	0.0058	0.0006	0.0000	0.0071
cor	--------	0.2181	0.0163	0.3104
S.E.	--------	0.0033	0.0012	0.0074
N	45	46	50	48
Oregon Inlet				
H_e	0.9340	0.8770	0.8660	0.9220
H_o	0.9790	0.6810	0.7610	0.7400
P	0.8493	**0.0000**	0.0523	**0.0000**
S.E.	0.0062	0.0000	0.0029	0.0000
cor	--------	**0.0000**	--------	**0.0001**
S.E.	--------	0.0000	--------	0.0001
N	47	47	46	50
Stumpy Point				
H_e	0.8940	0.8810	0.8230	0.9190
H_o	0.8800	0.5320	0.6000	0.6940
P	0.5670	**0.0000**	**0.0029**	**0.0000**
S.E.	0.0037	0.0000	0.0004	0.0000
cor	--------	0.0633	0.1493	**0.0001**
S.E.	--------	0.0018	0.0029	0.0000
N	50	47	50	49
Ocracoke Island				
H_e	0.9230	0.8800	0.8280	0.8640
H_o	0.9530	0.7020	0.6220	0.5810
P	0.5858	0.0200	**0.0000**	**0.0000**
S.E.	0.0081	0.0009	0.0000	0.0000
cor	--------	0.2297	**0.0096**	**0.0002**
S.E.	--------	0.0033	0.0011	0.0001
N	43	47	45	43

Table 4 cont. Microsatellite loci HWE P-values by location. H_e and H_o are percentage of heterozygous individuals expected and observed, respectively; P is the P-value from HWE; (cor) is the P-value from HWE using adjusted genotype frequencies from MICRO-CHECKER where appropriate (- indicates no evidence for null alleles); N is the number of individuals amplified with the respective loci. Significant P-values after sequential Bonferroni correction in bold.

	Cvi2i23	Cvi2j24	Cvi6	Cvi13
Cedar Island				
He	0.9020	0.8690	0.7880	0.9280
Ho	0.8780	0.7760	0.6810	0.6200
P	0.5117	0.2708	**0.0030**	**0.0000**
S.E.	0.0091	0.0031	0.0007	0.0000
cor	--------	--------	--------	**0.0000**
S.E.	--------	--------	--------	0.0000
N	49	49	47	50
Harker's Island				
H_e	0.9330	0.8790	0.8680	0.9260
H_o	0.8800	0.8570	0.6120	0.7400
P	0.3744	0.4985	**0.0001**	**0.0036**
S.E.	0.0086	0.0045	0.0001	0.0009
cor	--------	--------	**0.0082**	**0.0176**
S.E.	--------	--------	0.0007	0.0022
N	50	49	49	50
North Inlet, SC				
He	0.9110	0.8840	0.8110	0.9370
Ho	0.8330	0.7450	0.6040	0.7870
P	0.0369	0.0480	**0.0002**	**0.0136**
S.E.	0.0034	0.0015	0.0001	0.0017
cor	--------	0.0752	0.0142	**0.0134**
S.E.	--------	0.0019	0.0014	0.0016
N	48	47	48	47

Table 5. Cvi2j24 population genic (a) and genotypic (b) differentiation for adjusted data. P-value above diagonal, standard error below diagonal. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0157	0.7129	0.1318	0.4216	0.0029	0.0044	0.2061
WA	0.0007	-------	0.3087	0.5041	0.0248	0.2004	**0.0019**	0.2544
OI	0.0025	0.0033	-------	0.5336	0.3127	0.2287	0.0137	0.5655
SP	0.0023	0.0034	0.0035	-------	0.7037	0.4943	0.1852	0.9963
OC	0.0036	0.0011	0.0035	0.0034	-------	0.0912	0.5110	0.9562
CI	0.0002	0.0030	0.0031	0.0039	0.0020	-------	0.0296	0.5077
HI	0.0004	0.0002	0.0007	0.0028	0.0039	0.0011	-------	0.4585
SC	0.0029	0.0030	0.0035	0.0001	0.0010	0.0038	0.0038	-------

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0392	0.7603	0.1780	0.4955	0.0073	0.0062	0.2929
WA	0.0011	-------	0.3909	0.7037	0.0610	0.3114	0.0050	0.5369
OI	0.0024	0.0035	-------	0.5434	0.3764	0.2738	0.0188	0.6470
SP	0.0026	0.0027	0.0032	-------	0.7911	0.6405	0.2429	0.9983
OC	0.0030	0.0015	0.0033	0.0024	-------	0.1910	0.4091	0.9746
CI	0.0004	0.0030	0.0032	0.0032	0.0026	-------	0.0295	0.6490
HI	0.0003	0.0003	0.0007	0.0029	0.0035	0.0010	-------	0.4950
SC	0.0031	0.0034	0.0032	0.0001	0.0007	0.0032	0.0033	-------

Table 6. Cvi13 population genic (a) and genotypic (b) differentiation for adjusted data. P-value above diagonal, standard error below diagonal. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.7828	0.3316	0.1833	0.2509	0.6375	0.1462	0.1927
WA	0.0035	-------	0.6936	0.3943	0.2572	0.4946	0.1754	0.0547
OI	0.0043	0.0038	-------	0.7607	0.1750	0.4493	0.4442	0.0410
SP	0.0032	0.0044	0.0034	-------	0.2413	0.6575	0.1584	0.0912
OC	0.0043	0.0037	0.0032	0.0038	-------	0.4373	0.2819	0.0288
CI	0.0042	0.0045	0.0043	0.0039	0.0045	-------	0.6718	0.4702
HI	0.0032	0.0034	0.0046	0.0032	0.0045	0.0041	-------	0.1924
SC	0.0041	0.0019	0.0015	0.0025	0.0011	0.0049	0.0040	-------

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.9073	0.6493	0.5250	0.3581	0.9214	0.2753	0.3435
WA	0.0018	-------	0.9188	0.7258	0.3906	0.8522	0.3591	0.1427
OI	0.0032	0.0017	-------	0.9653	0.4530	0.9128	0.7459	0.1622
SP	0.0036	0.0032	0.0010	-------	0.4818	0.9576	0.4748	0.2248
OC	0.0034	0.0035	0.0036	0.0036	-------	0.7397	0.3859	0.0818
CI	0.0016	0.0023	0.0018	0.0011	0.0029	-------	0.9081	0.7392
HI	0.0034	0.0037	0.0033	0.0038	0.0037	0.0019	-------	0.2874
SC	0.0037	0.0027	0.0027	0.0030	0.0018	0.0032	0.0037	-------

Table 7. Cvi6 population genic (a) and genotypic (b) differentiation for adjusted data. P-value above diagonal, standard error below diagonal. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.8286	0.9490	0.2808	0.4632	0.4027	0.5071	0.7015
WA	0.0025	-------	0.6338	0.1314	0.7857	0.2121	0.6094	0.6487
OI	0.0012	0.0040	-------	0.2343	0.1852	0.3994	0.0999	0.3697
SP	0.0038	0.0027	0.0041	-------	0.0283	0.0112	**0.0018**	0.0177
OC	0.0043	0.0033	0.0035	0.0011	-------	0.4771	0.2725	0.0800
CI	0.0043	0.0034	0.0042	0.0007	0.0042	-------	0.0907	0.3744
HI	0.0038	0.0039	0.0024	0.0002	0.0035	0.0020	-------	0.0727
SC	0.0031	0.0036	0.0042	0.0009	0.0022	0.0042	0.0019	-------

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.9484	0.9583	0.3732	0.6752	0.6212	0.7083	0.9032
WA	0.0011	-------	0.8215	0.1837	0.9245	0.4439	0.7702	0.7784
OI	0.0009	0.0024	-------	0.2194	0.3559	0.5541	0.2107	0.4834
SP	0.0030	0.0024	0.0028	-------	0.0656	0.0316	0.0071	0.0362
OC	0.0031	0.0014	0.0036	0.0014	-------	0.7393	0.5411	0.1992
CI	0.0033	0.0034	0.0033	0.0009	0.0027	-------	0.2598	0.5065
HI	0.0029	0.0027	0.0027	0.0003	0.0034	0.0029	-------	0.1639
SC	0.0016	0.0025	0.0033	0.0010	0.0027	0.0034	0.0023	-------

Table 8. Cvi2i23 population genic (a) and genotypic (b) differentiation. P-value above diagonal, standard error below diagonal. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0856	0.3286	**0.0032**	0.1010	0.1485	0.0638	0.0581
WA	0.0079	-------	0.5141	0.0597	**0.0010**	0.1002	0.0071	0.0104
OI	0.0129	0.0146	-------	0.624	0.6229	0.4310	0.3564	0.3158
SP	0.0014	0.0049	0.0141	-------	**0.0004**	0.0625	**0.0005**	**0.0004**
OC	0.0082	0.0004	0.0154	0.0003	-------	0.0625	0.2245	0.6314
CI	0.0101	0.0071	0.0158	0.0067	0.0060	-------	**0.0027**	0.5780
HI	0.0064	0.0018	0.0150	0.0005	0.0129	0.0009	-------	0.0531
SC	0.0085	0.0017	0.0135	0.0003	0.0140	0.0146	0.0072	-------

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0786	0.3034	**0.0018**	0.1212	0.1772	0.0697	0.0601
WA	0.0018	-------	0.4672	0.0741	**0.0018**	0.1954	0.0100	0.0350
OI	0.0034	0.0038	-------	0.5946	0.5226	0.4206	0.3913	0.4524
SP	0.0002	0.0017	0.0035	-------	**0.0007**	0.0797	0.0024	0.0031
OC	0.0023	0.0002	0.0039	0.0001	-------	0.1012	0.2300	0.6777
CI	0.0026	0.0029	0.0037	0.0018	0.0021	-------	0.0108	0.6029
HI	0.0018	0.0005	0.0038	0.0002	0.0032	0.0006	-------	0.0735
SC	0.0016	0.0011	0.0038	0.0003	0.0032	0.0036	0.0018	-------

Table 9. Population genic (above diagonal) and genotypic (below diagonal) differentiation P-values (Fisher's method) across all microsatellite loci using adjusted data. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0823	0.7274	0.0040	0.2201	0.0178	0.0053	0.1046
WA	0.1617	-------	0.7213	0.1272	0.0026	0.1445	**0.0005**	0.0212
OI	0.8634	0.0089	-------	0.6772	0.2467	0.4148	0.0329	0.1546
SP	0.8550	0.5464	0.1146	-------	**0.0012**	0.0316	**0.0000**	**0.0007**
OC	0.0121	0.0064	0.3448	**0.0011**	-------	0.1042	0.3212	0.1067
CI	0.2786	0.0763	0.0179	0.4431	0.1345	-------	0.0034	0.6586
HI	0.7191	0.4842	0.0042	0.0138	0.4765	0.3414	-------	0.0393
SC	0.3907	0.6813	0.0936	0.2332	0.0059	0.8697	0.1179	-------

Table 10. a.)Population genotypic and genic differentiation P-values for microsatellite loci analyzed by 16s haplotype. Significant P-values after sequential Bonferroni correction in bold. b.) Population genotypic and genic differentiation P-values for global test (Fisher's method).

a.

	Genotypic	Genic
Cvi2i23	0.0993	0.0808
Cvi2j24	0.1797	0.2347
Cvi13	0.3576	0.1622
Cvi6	0.9885	0.9495

b.

	Genotypic	Genic
Global	0.2559	0.1664

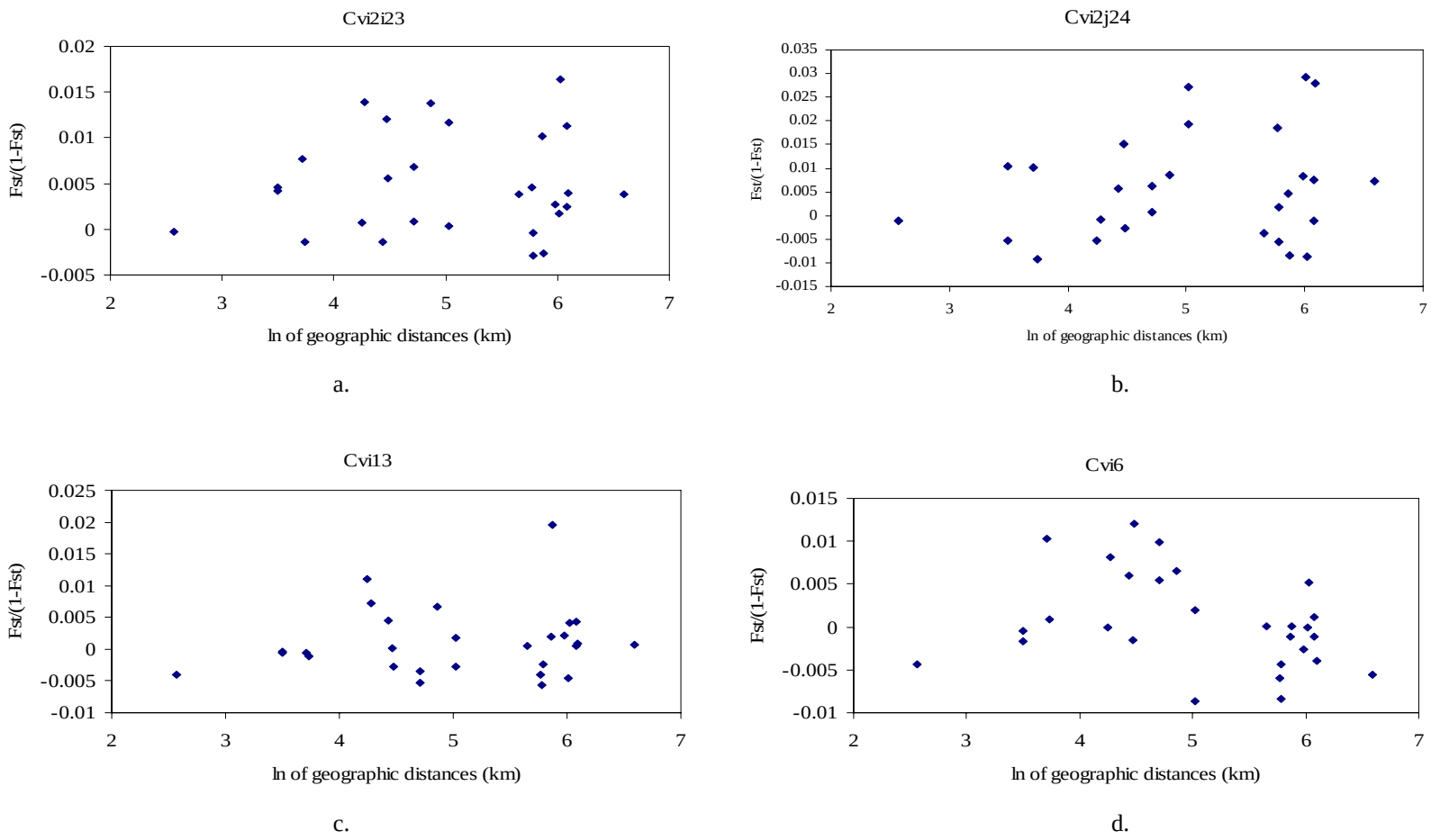

Figure 3. a.-d.) Microsatellite loci regression of population pairwise F_{ST} as computed in GENEPOP using Isolde. Test of IBD P-value follows: a.) 0.5056, b.) 0.2753, c.) 0.2797, d.) 0.8415.

Table 11. a.) Microsatellite AMOVA computed as standard data using pairwise differences, four loci, and 10100 permutations, with original data organized as one group in eight populations (by location). Φ_{ST} P-value = 0.0686. b.) Microsatellite AMOVA computed as standard data using pairwise differences, four loci, and 10100 permutations, with original data organized as one group in two populations (North Atlantic and South Atlantic) as determined by 16s sequencing. Φ_{ST} P-value = 0.2613

a.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	7	15.271	0.00535 Va	0.32
Within populations	784	1294.964	1.65174 Vb	99.68
Total	791	1310.235	1.65709	

b.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	1	2.172	0.00162 Va	0.10
Within populations	782	1289.990	1.64960 Vb	99.90
Total	783	1292.162	1.65123	

Table 12. Microsatellite loci (global) pairwise differences above diagonal. P-values below diagonal. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0031	-0.0034	0.0061	-0.0009	0.0081	0.0088	0.0038
WA	0.2847	-------	-0.0016	0.0011	0.0047	0.0024	0.0071	0.0065
OI	0.9749	0.8509	-------	-0.0007	0.0013	0.0040	0.0049	0.0027
SP	0.0932	0.5521	0.8027	-------	0.0032	0.0041	0.0108	0.0068
OC	0.7560	0.1462	0.4448	0.2935	-------	-0.0003	-0.0049	-0.0006
CI	0.0201	0.3272	0.1759	0.2000	0.6560	-------	0.0075	-0.0003
HI	0.0128	0.0313	0.0965	0.0023	0.9895	0.0201	-------	0.0018
SC	0.2072	0.0500	0.3019	0.0538	0.7096	0.7373	0.3965	-------

SNPs and RFLP

While all four loci were found to be polymorphic, not all populations were found to be variable. CvCHI was found to be fixed for the most common allele (A at nucleotide position 162) for both the Choptank River and the Wanchese populations, with little variation in frequency present in the remaining populations (Figure 4). Only Stumpy Point was found to deviate significantly from Hardy-Weinberg equilibrium (Table 13).

The most common allele for CvGP (C at nucleotide position175) varied in frequency for each population, though the frequency of the C allele was lowest in the Wanchese and Stumpy Point populations (88% and 89%, respectively), rather than at either of the extreme populations (Choptank River and South Carolina, 97% and 98% respectively). No populations were found to significantly deviate from Hardy-Weinberg equilibrium (Table 13).

Of the SNPs used in this study, CvRAN had largest frequency shift among populations at the most common allele (A at nucleotide position 100). This allele varied in frequency from a low of 35% at the northern most site to a high of 61% at the southern most site (Figure 4), though there were no significant deviations from Hardy-Weinberg equilibrium (Table 13).

A restriction fragment length polymorphism was observed in the Cv195 locus, and treated as a SNP for subsequent analysis. The most common allele (G) was present in moderate frequencies in all populations, with a low of 46% in the Choptank River population and a high of 64% in the Wanchese population (Figure 4). No statistically significant deviations from Hardy-Weinberg equilibrium were found for any populations using this locus.

Table 13. SNP Hardy-Weinberg Equilibrium P-values. (*) could not compute due to lack of variation. Significant values after sequential Bonferroni correction in bold.

	Locus			
	CvGP	CvRANB	CvCHI	Cv195
Choptank River				
P-value	1	0.5379	*	0.0843
S.E.	0	0.0027	--------	0.0023
Wanchese				
P-value	0.5439	0.0777	*	0.3491
S.E.	0.0022	0.0019	--------	0.0036
Oregon Inlet				
P-value	1	0.2495	0.0329	0.2307
S.E.	0	0.0033	0.0010	0.0031
Stumpy Point				
P-value	1	0.7680	**0.0001**	0.4000
S.E.	0	0.0018	0.0001	0.0033
Ocracoke Island				
P-value	1	0.3910	1	0.5510
S.E.	0	0.0037	0	0.0030
Cedar Island				
P-value	1	1	1	0.2558
S.E.	0	0	0	0.0036
Harker’s Island				
P-value	*	0.2524	1	0.7801
S.E.	--------	0.0033	0	0.0018
South Carolina				
P-value	1	0.7608	1	0.2500
S.E.	0	0.0016	0	0.0040

In a plot of the most common nucleotide at each position for the SNPs and RFLP loci (Figure 4), the CvGP and CvCHI loci most closely mirror each other, while the CvRANB SNP follows the same frequency trend across populations as the Cv195 RFLP. No linkage disequilibrium was found (P>0.05).

Significant population differentiation was found within the CvRANB locus in both the genic and the genotypic exact tests (Table 14). Across loci, the Choptank River population was also found to differ significantly from the Wanchese and Ocracoke Island collections, while Wanchese differed from Harker's Island (Table 18), in both exact tests. The remaining loci did not show any significant differences among populations analyzed by geography (Tables 15-17). Exact tests on SNP data analyzed by 16s haplotyped revealed no differences within individual loci, but the global tests both differed significantly (Table 19).

The possibility that the populations exhibited IBD was tested in the SNP and RFLP loci (ISOLDE GENEPOP). CvCHI had insufficient variation within the populations for this analysis. Of the three remaining loci, CvGP and Cv195 yielded regressions that were not significant, whereas CvRANB population regressions plotted against distance showed significance (Figure 5, P-value 0.0379). The three loci were also combined in a global SNP F_{ST} regression test that was also not significant.

SNP and RFLP data were grouped into locations as well as by 16s haplotype, similarly to the microsatellite data, for both AMOVA and population pairwise F_{ST} analysis (Table 20). Very little variation could be accounted for by either partition in the AMOVA analysis, though both results were significant. The population pairwise F_{ST} analyzed by 16s haplotype showed significant differences (F_{ST}: 0.0060, P-value: 0.0417), as did some of the populations partitioned by location (Table 21).

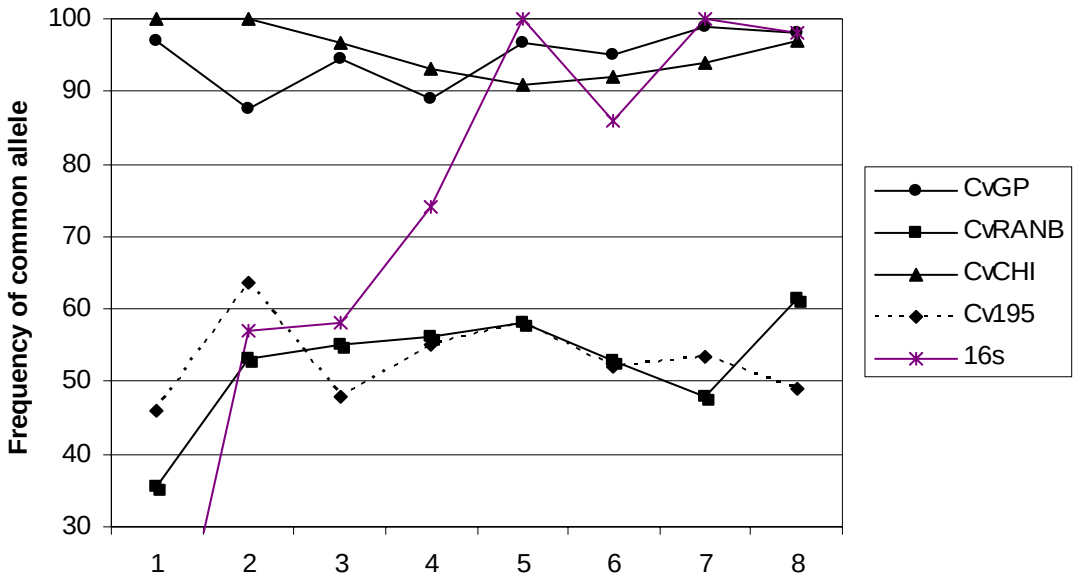

Figure 4. SNP and 16s most common allele frequency.1:Choptank River, 2:Wanchese, 3:Oregon Inlet, 4:Stumpy Point, 5:Ocracoke Island, 6:Cedar Island, 7:Harker's Island, 8:New Inlet.

Table 14. a.) CvRANB population genic (a) and genotypic (b) differentiation. Above diagonal: P-value, below diagonal: standard error. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0200	0.0060	0.0061	**0.0016**	0.0148	0.0873	**0.0005**
WA	0.0006	-------	0.8839	0.7687	0.5587	1	0.4779	0.3047
OI	0.0002	0.0006	-------	0.8866	0.7724	0.7787	0.3226	0.3842
SP	0.0003	0.0010	0.0006	-------	0.8858	0.6696	0.2552	0.5595
OC	0.0001	0.0017	0.0011	0.0006	-------	0.4770	0.1591	0.6631
CI	0.0005	0	0.0011	0.0014	0.0019	-------	0.5730	0.2535
HI	0.0012	0.0020	0.0020	0.0019	0.0017	0.0017	-------	0.0640
SC	0.0000	0.0019	0.0019	0.0017	0.0014	0.0019	0.0010	-------

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0149	0.0121	0.0086	0.0033	0.0189	0.0762	**0.0005**
WA	0.0003	-------	0.8831	0.7609	0.5486	1	0.4192	0.2594
OI	0.0003	0.0004	-------	0.8939	0.7895	0.7871	0.3248	0.3982
SP	0.0003	0.0008	0.0004	-------	0.8907	0.6752	0.2442	0.5575
OC	0.0001	0.0013	0.0008	0.0004	-------	0.4923	0.1522	0.6691
CI	0.0004	0	0.0008	0.0012	0.0014	-------	0.5517	0.2429
HI	0.0009	0.0013	0.0014	0.0014	0.0013	0.0013	-------	0.0468
SC	0.0000	0.0013	0.0015	0.0013	0.0011	0.0013	0.0007	-------

Table 15. CvGP population genic (a) and genotypic (b) differentiation. Above diagonal: P-value, below diagonal: standard error. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0152	0.4871	0.0426	1	0.7207	0.6227	1
WA	0.0003	-------	0.1263	0.8221	0.0288	0.0767	0.0026	0.0054
OI	0.0008	0.0009	-------	0.2829	0.4943	1	0.2125	0.2671
SP	0.0005	0.0005	0.0011	-------	0.0471	0.1784	0.0096	0.0155
OC	0	0.0004	0.0008	0.0005	-------	0.7225	0.6206	0.6769
CI	0.0005	0.0007	0	0.0009	0.0005	-------	0.2155	0.4448
HI	0.0005	0.0001	0.0007	0.0002	0.0005	0.0007	-------	1
SC	0	0.0001	0.0008	0.0002	0.0005	0.0008	0	-------

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0179	0.4744	0.0352	1	0.7143	0.6191	1
WA	0.0004	-------	0.1340	0.8203	0.0322	0.0825	0.0038	0.0061
OI	0.0008	0.0011	-------	0.2576	0.4839	1	0.2047	0.2592
SP	0.0004	0.0007	0.0011	-------	0.0401	0.1592	0.0075	0.0130
OC	0	0.0005	0.0008	0.0004	-------	0.7158	0.6171	0.6748
CI	0.0005	0.0009	0	0.0009	0.0005	-------	0.2076	0.4371
HI	0.0005	0.0001	0.0007	0.0001	0.0005	0.0007	-------	1
SC	0	0.0002	0.0008	0.0002	0.0005	0.0008	0	-------

Table 16. CvCHI population genic (a) and genotypic (b) differentiation. Above diagonal: P-value, below diagonal: standard error. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	*	0.1186	0.0142	0.0033	0.0071	0.0296	0.2468
WA	*	-------	0.1117	0.0143	0.0032	0.0069	0.0290	0.1184
OI	0.0005	0.0005	-------	0.3331	0.1352	0.2160	0.4985	1
SP	0.0002	0.0002	0.0009	-------	0.7954	1	1	0.3325
OC	0.0001	0.0001	0.0008	0.0005	-------	1	0.5926	0.1341
CI	0.0001	0.0001	0.0009	0	0	-------	0.7827	0.2125
HI	0.0003	0.0003	0.0008	0	0.0009	0.0005	-------	0.4977
SC	0.0006	0.0005	0	0.0010	0.0008	0.0009	0.0009	-------

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	*	0.2410	0.1172	0.0026	0.0057	0.0272	0.2421
WA	*	-------	0.2331	0.1179	0.0026	0.0059	0.0266	0.1177
OI	0.0008	0.0008	-------	0.4888	0.1606	0.2517	0.5490	1
SP	0.0007	0.0007	0.0010	-------	0.8256	1	1	0.4532
OC	0.0001	0.0000	0.0011	0.0005	-------	1	0.5785	0.1210
CI	0.0001	0.0001	0.0011	0	0	-------	0.7740	0.2003
HI	0.0003	0.0002	0.0009	0	0.0009	0.0009	-------	0.4876
SC	0.0006	0.0005	0	0.0010	0.0007	0.0008	0.0008	-------

Table 17. Cv195 population genic (a) and genotypic (b) differentiation. Above diagonal: P-value, below diagonal: standard error. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0142	0.8845	0.1994	0.1166	0.4719	0.7761	0.7701
WA	0.0005	-------	0.0406	0.3039	0.4655	0.1447	0.0315	0.0560
OI	0.0006	0.0008	-------	0.3808	0.1946	0.6628	1	1
SP	0.0019	0.0019	0.0020	-------	0.7723	0.7721	0.3183	0.4684
OC	0.0014	0.0019	0.0018	0.0011	-------	0.4713	0.1600	0.2453
CI	0.0019	0.0015	0.0014	0.0011	0.0019	-------	0.6703	0.7696
HI	0.0011	0.0007	0	0.0020	0.0018	0.0015	-------	1
SC	0.0011	0.0010	0	0.0019	0.0018	0.0010	0	-------

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	-------	0.0122	0.8683	0.1539	0.0815	0.4613	0.7655	0.7678
WA	0.0003	-------	0.0379	0.3086	0.4691	0.1779	0.0421	0.0800
OI	0.0004	0.0006	-------	0.3387	0.1542	0.6613	1	1
SP	0.0011	0.0015	0.0014	-------	0.7585	0.7761	0.3140	0.4695
OC	0.0008	0.0015	0.0011	0.0008	-------	0.4784	0.1514	0.2523
CI	0.0014	0.0014	0.0011	0.0008	0.0014	-------	0.6850	0.7881
HI	0.0008	0.0007	0	0.0015	0.0012	0.0011	-------	1
SC	0.0008	0.0009	0	0.0015	0.0014	0.0008	0	-------

Table 18. SNP and RFLP global genic (a) and genotypic (b) differentiation. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

a.

	CH	WA	OI	SP	OC	CI	HI	SC
CH		**0.0003**	0.0399	**0.0004**	**0.0003**	0.0087	0.0973	0.0196
WA			0.0558	0.1609	0.0065	0.0151	**0.0006**	0.0035
OI				0.5480	0.3256	0.8206	0.5636	0.8041
SP					0.5019	0.7822	0.0743	0.1048
OC						0.8884	0.3142	0.3926
CI							0.7060	0.4349
HI								0.5493
SC								

b.

	CH	WA	OI	SP	OC	CI	HI	SC
CH		**0.0003**	0.0934	0.0020	**0.0003**	0.0083	0.0845	0.0170
WA			0.0880	0.4723	0.0058	0.0161	**0.0008**	0.0048
OI				0.5857	0.3173	0.8496	0.5756	0.8056
SP					0.4689	0.7619	0.0604	0.1123
OC						0.8933	0.2958	0.3819
CI							0.6898	0.4158
HI								0.4714
SC								

Table 19. Genotypic and genic population differentiation for SNP loci analyzed by 16s haplotype. Significant P-values after sequential Bonferroni correction in bold.

Locus	Genotypic	Genic
CvRANB	0.0160	0.0165
CvGP	0.5990	0.5966
CvCHI	0.0367	0.0328
Cv195	0.5192	0.4687
Global	**0.0279**	**0.0245**

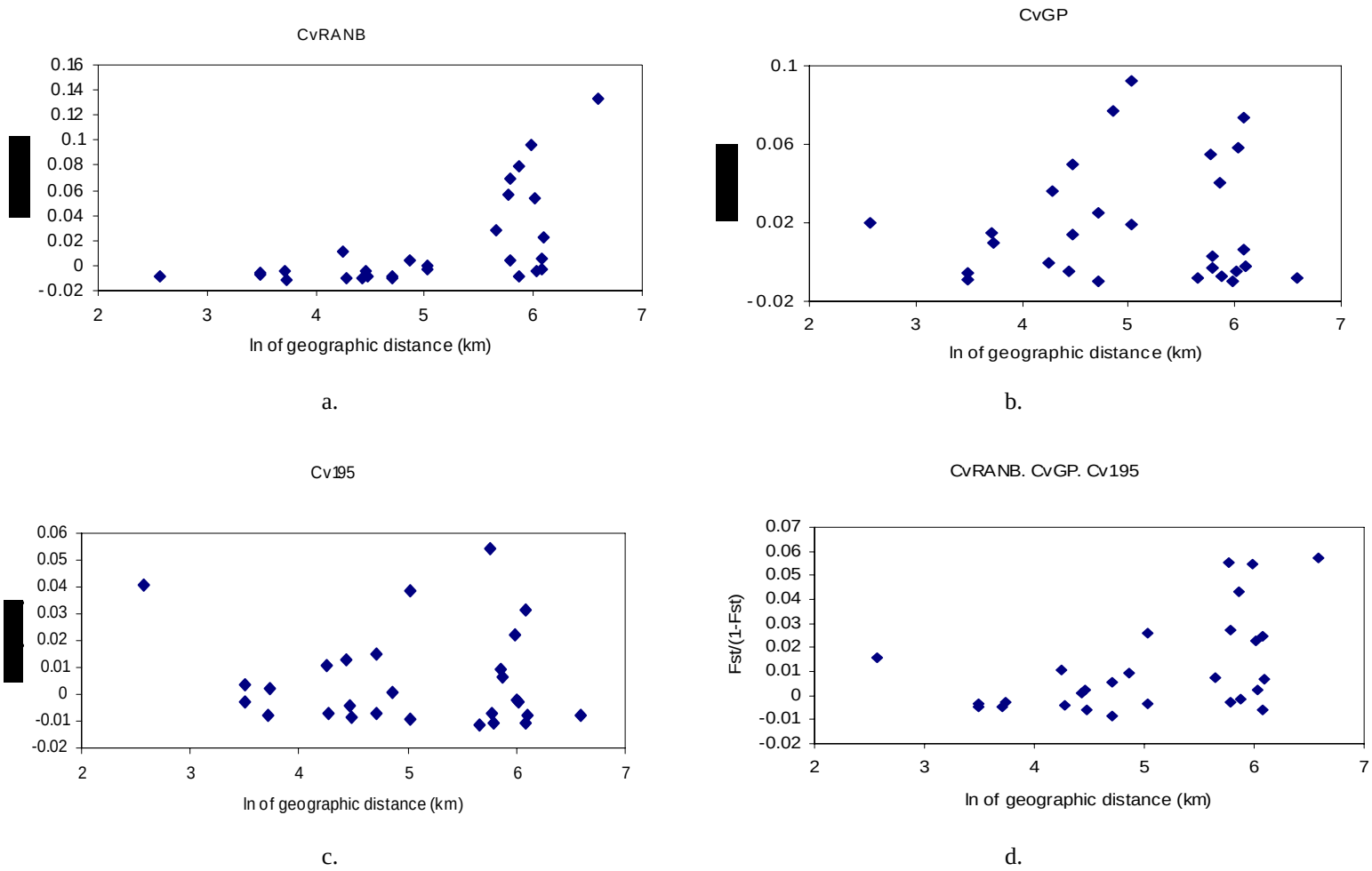

Figure 5. a.-d.) SNP loci regression of population pairwise F_{ST} as computed in GENEPOP using Isolde. Test of IBD P-value follows: a.) **0.0379**, b.) 0.5464, c.) 0.7348, d.) 0.0790.

Table 20. a.) SNP and RFLP AMOVA (Arlequin), Data grouped by location. Distance method: pairwise differences. Significance tests: 10100 permutations, Φ_{ST}: 0.0110, P-value = **0.0049** +- 0.0007
b.) SNP and RFLP AMOVA (Arlequin). Data grouped by 16s haplotype. Distance method: pairwise differences. Significance tests: 10100 permutations, Φ_{ST}: 0.0060, P-value = **0.0391**+- 0.0021

a.

Source of Variation	d.f.	Sum of Squares	Variance components	% of Variation
Among Populations	7	8.158	0.00615 Va	1.10
Within Populations	790	436.264	0.55223 Vb	98.90
Total	797	444.422	0.55838	100.00

b.

Source of Variation	d.f.	Sum of Squares	Variance Components	% of Variation
Among Populations	1	1.648	0.00338 Va	.60
Within Populations	786	437.807	0.55701 Vb	99.4
Total	787	439.454	0.56039	100.00

Table 21. SNP and RFLP population pairwise FST's computed in Arlequin. Data grouped by location. Above diagonal: pairwise differences. Below diagonal: FST P- values. Number of permutations: 10100. Significant P-values after sequential Bonferroni correction in bold. CH = Choptank River, WA = Wanchese, OI = Oregon Inlet, SP = Stumpy Point, OC = Ocracoke Island, CI = Cedar Island, HI = Harker's Island, and SC = New Inlet.

	CH	WA	OI	SP	OC	CI	HI	SC
CR	-------	0.0546	0.0279	0.0436	0.0504	0.0222	0.0032	0 .0518
WA	**0.0013**	-------	0.0150	0.0002	0.0068	0.0119	0.0192	0.0282
OI	0.0156	0.0641	-------	-0.0046	-0.0003	-0.0092	-0.0072	-0.0032
SP	**0.0019**	0.3802	0.6398	-------	-0.0035	-0.0034	0.0043	0.0082
OC	**0.0019**	0.1593	0.3893	0.5859	-------	-0.0057	0.0076	0.0018
CI	0.0333	0.0944	0.9368	0.6030	0.7327	-------	-0.0086	-0.0004
HI	0.2414	0.0352	0.7929	0.2250	0.1458	0.9320	-------	0.0052
SC	0.0022	0.0177	0.5201	0.1523	0.3039	0.4022	0.2123	-------

DISCUSSION

In general, there was an absence of significant structure across the nuclear genome loci tested, especially for the microsatellite loci. A check for Hardy-Weinberg Equilibrium using microsatellite data revealed heterozygote deficiencies in all of the populations in 3 of the microsatellite loci. These can best be explained by the presence of null alleles, despite attempts to correct for genotype frequencies. Previous studies identified polymorphisms in the flanking region (Cvi2i23: 2.1%, Cvi2j24: 4.5%) which could contribute to null alleles by preventing the annealing of primers. Heterozygote deficiencies have been reported for both Cvi13 and Cvi6 (Reece et al, 2004; Brown et al, 2000), which could be indicative of null alleles. Using a multi-locus approach, MICRO-CHECKER can discriminate between the distinctive genotypic signatures the null alleles would present and, if detected, correct the genotype frequencies in those populations (Van Oosterhout et al, 2004). Null alleles were detected and corrected for in several populations within Cvi2j24, Cvi13 and Cvi6. Null alleles could still be present in the data set, as the methods used to correct for genotypes in this study assume a single null allele, whereas multiple primer site polymorphism could result in a variety of different null alleles being present in varying frequencies (Rose, et al. 2006). Genotype corrections resulted in a decrease in the number of populations out of Hardy-Weinberg Equilibrium at each locus. However, all 8 populations still deviated significantly at 1 or more loci (Table 4), despite genotypic corrections. Null alleles then, given their suspected high frequency could be the cause of the significant differences seen in the microsatellite exact tests.

Exact tests of individual SNP/RFLP loci after sequential Bonferroni correction revealed only one significant difference, and that was at the CvRANB locus. However, it should be noted

that had such a conservative method of correction not been used, the Choptank River population within the CvRANB locus would be significantly different from all other populations with the exception of Harker's Island. A global test combining the SNPs and the RFLP also suggests genotypic differentiation between populations.

Tests of IBD established a pattern for the differences seen in the CvRANB exact test. A regression of computed F_{ST} values revealed significant differences in relation to geographic distances, but only in the CvRANB locus (Figure 4). IBD is often modeled in a stepping stone fashion, though it has been shown that local recruitment combined with some long distance dispersal can reveal this same pattern (Palumbi, 2003). This type of neutral gene flow should be seen across multiple loci, however, as it would affect the entire genome, rather than a single locus.

The presence of population structure in the SNP/RFLP loci is supported by both AMOVAs, though more of the variation is accounted for when samples are grouped by location rather than by 16s haplotype. The genotypic differences seen in the exact tests in addition to the IBD tests suggest that either the CvRANB locus or the Choptank River population is driving the significant Φ_{ST}s seen in the AMOVA. Subsequent reanalysis in which the CvRANB locus was removed from the 8 populations still yielded a significant Φ_{ST} value, whereas a reanalysis in which the Choptank River population was removed across all 4 loci (grouped by both location as well as by 16s haplotype) revealed no significant structure. When the Choptank River population was grouped by itself and the remaining 7 populations were grouped together in an AMOVA involving all 4 loci, the among population variation accounted for 3.4% of the total, and the resultant Φ_{ST} (0.0339) was significant ($P = 0.0048$). Pairwise differences also show significant

deviations in both population (Choptank River vs. other populations) and 16s grouping, though this should be interpreted with caution, due to the inclusion of the Choptank River population.

The mtDNA cline seen in *Crassostrea virginica* off the coast of Florida occurs over an ~ 340 km range (Reeb and Avise, 1992), with an abrupt ~20 km change in mtDNA frequencies that separates oysters into Atlantic and Gulf assemblages and is generally supported by both allozyme and nDNA analysis in spite of their dispersal potential and a lack of obvious genetic barriers (Hare and Avise, 1996; but see McDonald et al., 1996). Cytonuclear disequilibria, vicariant separation with secondary contact, and balancing selection on nuclear genomic proteins have all been suggested as possible explanations for the maintenance of this cline (Karl and Avise, 1992; Hare and Avise, 1998).

The Gulf and Atlantic assemblages are reciprocally monophyletic for their respective mtDNA RFLP variants (Hare and Avise, 1998). Due to effective population size differences between the nuclear genome and the maternally inherited mtDNA genome, Palumbi (2001) suggested that it would take nDNA 3 times as long as mtDNA to coalesce to reciprocal monophyly. Hare (1998) reanalyzed the fragment approach Reeb and Avise (1990) used, excluding all of the data subject to artifact and found that, while remaining reciprocally monophyletic, not enough time has passed for the nDNA to sort into distinct lineages.

So why was structure seen in the mitochondrial genome and not the nuclear genome within Pamlico Sound? Karl and Avise (1992) invoked balancing selection as a factor in the discrepancy between the combined mtDNA and scnDNA and Buroker's (1983) allozyme frequencies, although it has been shown that those data sets are not as much in disagreement as previously thought. There is no need to invoke balancing selection in this case, however, as there is no evidence of a lack of gene flow.

An alternative to balancing selection could be directional selection. The Choptank River population appears to be driving the majority of the significant structure seen in both the CvRANB exact tests as well as within the combined SNP/RFLP AMOVAs. If certain genes increase an oyster's fitness in different abiotic regimes, those genes can then be selected for, and this could be reflected in neutral loci. This fails to explain the homogeneity seen in the allozyme data, however (Buroker, 1983).

Assuming there is no barrier to gene flow, what we are seeing in the mtDNA cline is either introgression or admixture of the northern and southern 16s haplotype. While the SNP/RFLP AMOVA organized by 16s was significant, more of the variation within populations was accounted for when populations were analyzed by location, which suggests that northern and southern oysters are not reproductively isolated. Introgression of nDNA, then, could account for the overall lack of signal seen in the nuclear genome within Pamlico Sound, but fails to explain the maintenance of the mtDNA cline.

Sweepstakes events (extreme variance in reproductive success) could account for the deviations from Hardy-Weinberg seen in the microsatellites, as well as the genotypic differences seen in the exact tests (Hedgecock, 1994). This would seem unlikely, however, as deviations from Hardy-Weinberg were not seen in the Cvi2i23 locus nor the SNP/RFLP analysis (with the single exception of the Stumpy Point population). A sweepstakes event would result in a general deficit of heterozygosity across loci and this should be reflected in multi-locus comparisons.

It seems likely that purely neutral mechanisms could explain the significant deviations seen in this study. Finnelli and Wethey (2003) monitered oyster larvae from 15-21 days post-fertilization and observed size-dependent behavior that is suggestive of induced vertical distribution. Larval vertical distribution simulations in response to salinity and temperature

regimes coupled with horizontal flow have also been suggested to play an important part in larval retention or distribution (Dekshenieks et al., 1996), and oyster larvae have been seen to maintain their position in a water column as well as move upstream in some trap-like tributaries in the Chesapeake Bay (Southworth and Mann, 1998). It could be that, rather than directional selection, local-retention within the Choptank River has reduced gene flow within that population, and that this potentially inbred population is the reason behind the exact tests, IBD and AMOVA differences seen in this data.

Loci analyzed by location shows some population differentiation, mostly within Pamlico Sound, while structure is found only when the Choptank River population is included in the analysis. This suggests that there could be a lack of gene flow between Pamlico Sound populations and the Choptank River population. Alternatively, anthropogenic effects could be influencing the analysis. Chesapeake Bay oysters have experienced extreme pressures in past decades, and ongoing efforts to restore the flagging eastern oyster to prominence have been underway for some time. Perhaps drift acting on small populations has reduced enough of the genetic variation within the Chesapeake Bay to register a genetic signature of differentiation in the face of ongoing gene flow. Or introgression on Gulf nDNA from Louisiana oysters that have been planted there could be causing the deviations seen, as the Gulf mtDNA haplotype has been detected in newly recruited spat (Milbury et al., 2004).

Loci analyzed by 16s haplotype yielded results similar to the data analyzed by location. With the Choptank River population removed (all northern 16s haplotypes), no structure was evident, which suggests that the two genomes are not concordant with respect to the mtDNA cline. If the two assemblages are reproductively isolated, it's possible that the nDNA simply hasn't had sufficient time to sort into respective lineages. However, there are no known

mechanisms of within species reproductive isolation that would be sufficient to maintain the cline. Also, due to fecundity and hermaphroditic life history, the effective population size, and thus the sorting time, between mtDNA, microsatellites and SNPs may not be all that different.

Either the mtDNA is showing a signature of historic separation which did not last long enough to sort nDNA but should also erode under contemporary gene flow, or selection is maintaining mtDNA is the face of gene flow. Genetic analysis of larvae at the extremes of the cline followed by sampling and genetic analysis of first year cohorts could potentially allow insight into the extent of gene flow across the cline. If oysters exhibiting a northern haplotype are being transported into an area containing predominantly southern haplotypes (and vice versa), one can assess mortality, and hence selection, by examining the haplotypes in the survivors. If a genetic analysis of oyster larvae does not show transport of the minority haplotype into the majorities range, one can assess gene flow into a region.

LITERATURE CITED

Avise, J.C. 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*. 63:62-76.

Avise, J.C. 2004. *Molecular Markers, Natural History, and Evolution*. Sinauer Associates, Inc., Sunderland, MA, p. 419.

Barber, J. B., S.E. Ford, and R.N. Wargo. 1991. Genetic variation in the timing of gonadal maturation and spawning of the Eastern oyster, *Crassostrea virginica* (Gmelin). *Biol. Bull*. 181: 216-221.

Brown, B.L., D.E. Franklin, P.M. Gafney, M. Hong, D. Dendantos, and I. Kornfield. 2000. Characterization of microsatellite loci in the eastern oyster, *Crassostrea virginica*. *Molecular Ecology*. 9: 2217-2219.

Brownstein, M.J., J.D. Carpten, and J.R. Smith. 1996. Modulation of Non-Templated Nucleotide Addition by *Taq* DNA Polymerase: Primer Modifications that Facilitate Genotyping. *BioTechniques*. 20: 1004-1010.

Buroker, N.E. 1983. Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. *Marine Biology*. 75: 99-112.

Cunningham, C.W. and T.M. Collins. 1994. Developing model systems for molecular biogeography: Vicariance and interchange in marine invertebrates. In: *Molecular Ecology and Evolution: Approaches and Applications* (Schierwater, B., Streit, B., Wagner, G. P., DeSalle, R., eds.). Birkhauser Verlag, Switzerland, pp. 405-433.

Dekshenieks, M.M., E.E. Hofmann, J.M. Klinck, and E.N. Powell. 1996. Modeling the vertical distribution of oyster larvae in response to environmental conditions. *Mar. Ecol. Prog. Ser*. 136: 97-110.

Dittman, D.E., S.E. Ford, and H.H. Haskin. 1998. Growth patterns in oysters, *Crassostrea virginica*, from different estuaries. *Marine Biology*. 132: 461-469.

Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*. 131: 479-491.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*. 1:47-50.

Finelli, C.M. and D.S. Wethey. 2003. Behavior of oyster (*Crassostrea virginica*) larvae in flume boundary layer flows. *Marine Biology*. 143: 703-711.

Glaubitz, J.C. 2004. CONVERT: a user friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*. 4: 309-310.

Hare, M.P. and J.C. Avise. 1996. Molecular genetic analysis of a stepped multilocus cline in the American Oyster (*Crassostrea virginica*). *Evolution*. 50(6): 2305-3215.

Hare, M.P. and J.C. Avise. 1998. Population structure in American oysters as inferred by nuclear gene genealogies. *Molecular Biology and Evolution*. 15(2): 119-128.

Hedgecock, D. 1982. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Science*. 39:550-564.

Hedgecock, D. 1994. Does variance in reproductive success limit effective population size of marine organisms? In. *Genetics and evolution of aquatic organisms* (Beaumont AR, ed). London: Chapman and Hall; 122-134.

Hedrick, P.W. 1999. Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation. *Evolution*. 53(2): 313-318.

Hillman, R.E. 1964. Chromatographic evidence of intraspecific genetic differences in the Eastern oyster, *Crassostrea virginica*. *Systematic Biology*. 13:12-18.

Karl, S.A. and J.C. Avise. 1992. Balancing selection at allozyme loci in oysters: Implications from nuclear RFLPs. *Science*. 256: 100-102.

Loosanoff, V.L. and C.A. Nomejko. 1951. Existence of physiologically-different races of oysters, *Crassostrea virginica*. *Biological Bulletin*. 101: 151-156.

McDonald, J.H., B.C. Verelli, and L.B. Geyer. 1996. Lack of geographic variation in anonymous nuclear polymorphisms in the American oyster, *Crassostrea virginica*. *Mol. Biol. Evol*. 13(8): 1114-1118.

Palumbi, S.R. 1999. All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences of the United States of America*. 96(22): 12632-12637.

Posada, D. and K.A. Crandall. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics*. 14: 817-818.

Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*. 86: 248-249.

Reeb, C.A. and J.C. Avise. 1990. A genetic discontinuity in a continuously distributed species: Mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics*. 124: 397-406.

Reece, K.S., W.L Ribeiro, P.M Gaffney, R.B. Carnegie, and S.K. Allen, Jr. 2004. Microsatellite Marker Development and Analysis in the Eastern Oyster (*Crassostrea virginica*): Conformation of Null Alleles and Non-Mendelian Segregation Ratios. *Journal of Heredity*. 95(4): 346-352.

Rezak, R., T.J. Bright, and D.W. McGrail. 1985. *Reefs and banks of the Northwestern Gulf of Mexico.* John Wiley & Sons. New York.

Rice, W.R. 1989. Analyzing Tables of Statistical Tests. *Evolution*. 43(1): 223-225.

Rose, C.G., K.T. Paynter, and M.P. Hare. 2006. Isolation by Distance in the Eastern Oyster, *Crassostrea virginica*, in Chesapeake Bay. *Journal of Heredity*. 97(2): 158-170.

Sackett, R.E. 2002. Characterization of North Carolina *Crassostrea virginica* population structure based on mtDNA haplotype variation. M. Sc. Thesis, University of North Carolina at Wilmington, NC.

Southworth, M. and R. Mann. 1998. Oyster Reef Broodstock in the Great Wicomico River, Virginia. *Journal of Shellfish Research*. 17(4): 1101-1114.

Van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*. 4: 535-538.

Weir, B.S. and W.G. Hill. 2002. Estimating F-statistics. *Annual Review of Genetics*. 36:721-750.

Wakefield, J.R. 1996. Sequence variation on the mitochondrial large subunit (16s) ribosomal gene of the American oyster, *Crassostrea virginica*. M. Sc. Thesis, University of Delaware, DE.

APPENDIX

Appendix A. Loci/Primer combinations examined during the initial phases of this study

Locus/Primers	Source	Difficulty
OYARGKI/ F1-R1	Gaffney, P.	Lack of Consistent Amplification
OYARGK1/ F1-R2	Gaffney, P.	Lack of Consistent Amplification
OYARGK2/ F1-R3	Gaffney, P.	Lack of Polymorphisms
OYPES	Gaffney, P.	Lack of Consistent Polymorphisms
53U23-203L21		Lack of Amplification
Acetycholine (197U21-274L23)		Lack of diagnostic sites for RFLP
Coactosin (109U21-435L24)		Some polymorphisms, lack of diagnostic sites for RFLP
Cam2F/R		Plenty of polymorphism, lack of diagnostic sites for RFLP
CvL1	McDonald et al. (1996)	Plenty of polymorphism, lack of diagnostic sites for RFLP, multiple sized amplifying fragments
LTRS-F1/R1 CvUSB-F/R		Non-specific hybridization Lack of polymorphism
Cv7.7-L/R	Hare, Karl and Avise (1996)	Seems strong?
Cvi2i10		Would not amplify
Cv233-L/R	Hu and Foltz	Little polymorphism
Cv19-L/R	Hu and Foltz	Large amplicon, >1k

Appendix B. Latitude and Longitude used to determine distance (in kilometers) between populations.

	CH	WA	OI	SP	OC	CI	HI	SC
CH	384013N 761330W							
WA	319	355034N 753819W						
OI	324	013	354605N 753126W					
SP	352	033	042	354526N 755923W				
OC	396	087	084	072	350653N 755852W			
CI	408	111	111	088	033	350030N 761858W		
HI	443	152	152	129	070	041	344154N 763335W	
SC	728	437	437	414	355	326	285	331939N 790958W

Appendix C. Consensus sequence for 3 SNP loci and 1 scNDA. Targeted sites in bold.

CVIGP

GCAAGCCATTCCGAAGATACAGCCGCCATTATATCATGACAAATATACAGAGCGACATCAAGAAACTTATAGACCAAT
TCCAAGTTAAATAATCAAAATTATCTACCGTCATATTTTGCTGTTGTTTGAATCTTGTTCAATAAGTGTAACAAGTAAGT
GTATATGTGAAAAGGT**S**ATTTGCTTTAATGATTATTCTTCACAAAGCTCCTTATACCCGAGACGGGAGGATATGATGTG
CACGCAA

CVRAN

AAATGTTCCCAACTGGCATAGAGATCTGGTCCGCGTGTGTGAAAACATCCCCATTGTGTTATGTGGAAACAAAGTCGAC
ATCAAGGATCGCAAAGTTAA**R**GCTAAAGCCATCGTGTTTCACCGGAAAAAGAACTTGCAGTACTACGACATAAGCGCC
AAGAGTAACTACAACTTCGAGAAGCCTTTCCTYTGGTTAGCTAGGAAATTGGTGGGAG

CVCHI
CGGCAGAGTACTGGCACCAGAAGGGCGCYCCYAGGGASAAACTGATCATCGGYCTGGCTACTTAYGGRCGGAGCTTCA
CCCTAGTGGACAGTTCTCAGCACGGGGTCGGAGCCCCKGTGTCAGGGGCCGGGCTAGCTGGGCCCTATACAAGGGAGA
AAGGT**K**TCYTGTCCTATTACGAGGTATTGCAAAGTRWAAAWTGRCCTTTCRTCGGTTACAGTAGATTAATAACTTAATA
CTATAGTTTAHAGRAAACAGTTTCCGGGAGCAATAACG

Cv195

AGAATAGTAATAAATGACTGGGAAAAGACGAAGTCTAAGTTTCCAGAGTTACCGGTCGGGGGGTCGTTTCAAATACTT
CATCCAAGAATGGAAATTGATAACTCAAGACAATAGGGTCCTGTCAATACTCAAAGAGCGGTACAAAATATAATCTAT
TCAAAAACCACAATTTCTCA**R**GGTGAAAAAGACATC