Characterization of EST derived SSRs from the bay scallop, *Argopecten irradians*

Steven Roberts, Christina Romano, and Gabriele Gerlach

Marine Biological Laboratory, Woods Hole, MA 02543

Abstract

Interest in bay scallop conservation has resulted in organized stock enhancement efforts and increased attention to fisheries management issues. Genetic markers can facilitate the monitoring of enhancement efforts, characterization of wild populations, and optimize hatchery practices. We have identified eight polymorphic simple sequence repeat markers including one dinucleotide repeat, six trinucleotide repeats and one compound dinucleotide repeat, in expressed sequence tags generated from multiple bay scallop cDNA libraries. The numbers of alleles range from 2 – 5. The expected and observed heterozygosities range from 0.093-0.720 and 0.095-0.600, respectively.

The bay scallop (*Argopecten irradians*) is found in the western North Atlantic along the majority of the United States coast. The bay scallop is a hermaphroditic bivalve mollusk with an enlarged, single adductor muscle that is able to quickly clasp both valves together for propulsion. Commercial landings of bay scallops in the United States peaked at over 1200 metric tons in 1982, corresponding to more than 11 million dollars (US) of revenue (NMFS 2004). Over the past twenty years populations have experienced significant declines that have been attributed to several factors including habitat loss and pressure from commercial and recreational fisheries. One approach to enhance bay scallop populations involves spawning wild broodstock in a hatchery and releasing

offspring into the wild (Arnold 2001). The aim of this work was to identify genetic markers that could be used in evaluating stock enhancement efforts. These same markers will also aid in characterizing population dynamics of native populations and assist aquaculture operations efficiently manage stocks.

In order to identify genetic markers, bay scallop expressed sequence tags (ESTs) were screened for simple sequence repeats (SSRs). This approach has been used successfully across several taxa, and is prevalent in the field of plant genomics (i.e. Kantety et al 2002). To date, we have generated and submitted 2089 cDNAs to NCBI's EST database from the bay scallop (Roberts et al. unpublished, Roberts and Goetz 2003). These ESTs are from developing larvae as well as from adductor muscle and gonad tissue. Simple sequence repeats were identified using the Simple Sequence Identification Tool (http://www.gramene.org/db/searches/ssrtool) (Temnykh et al. 2001). Primers flanking SSRs with a minimum of 5 tandem repeats and 14 bps were designed using Macvector 7.2 (Accelrys) or by visual inspection.

DNA was isolated from mantle tissue of bay scallops collected from Cape Cod, MA utilizing a Chelex (BioRad) extraction technique (Walsh et al, 1991). The PCR was carried out under the following conditions: initial denaturation for 3 min at 94^{o}C followed by 30-40 cycles of 1 min at 94^{o}C, 1 min at 52-58 ^{o}C according to the specific primer set used (Table 1) and 1 min at 72^{o}C with a final extension step of 7 min at 72^{o}C. PCRs were carried out in a total volume of 18 µl containing 0.2 µM each of forward and reverse primers, 0.2 µM dNTPs (Invitrogen), 1.8 µl 10X AmpliTaq Buffer containing 15 mM $MgCl_2$ (Applied Biosystems), 0.36 U AmpliTaq DNA Polymerase (Applied Biosystems), and 0.8 µl DNA template.

Electrophoresis of the PCR fragments was performed using the SEA 2000TM advanced submerged gel electrophoresis apparatus (Elchrom Scientific AG, Switzerland). Products and a 10 bp DNA Ladder (Invitrogen) were separated on Spreadex EL-400 or EL-600 S-100 gels (Elchrom Scientific AG, Switzerland) at 95-150 V and 998 mA for 30-45 minutes at 55^{o}C; and visualized by staining with SYBR Gold. Alleles were scored manually by visual inspection. The data were first analyzed using Micro-checker (van Oosterhout, et al. 2004) for the presence of genotyping errors. Expected and observed heterozygosities, and Fis values were calculated using FSTAT version 2.93 (Goudet 1995) (Table 1). Pairwise tests for linkage disequilibrium were also performed using FSTAT. Sequencial Bonferroni corrections were applied for all multiple tests.

Primers were designed for 29 SSRs, of which 8 were not easily amplified, 13 were monomorphic, and 8 were polymorphic. The eight polymorphic SSRs identified include one dinucleotide repeat, six trinucleotide repeats and one compound dinucleotide repeat. The numbers of alleles range from 2 – 5. The expected and observed heterozygosities range from 0.093-0.720 and 0.095-0.600, respectively. The reduced heterozygosity and departure from Hardy-Weinberg equilibrium at locus GL23, suggests the presence of null alleles. Further, tests for null alleles using Micro-checker were only significant at GL23. There is no linkage disequilibrium between any loci, including two pairs of loci located within single ESTs (GP340:GP63 and C1831:C1832). The p-values resulting from tests of linkage disequilibrium for G340-GP63 and C1831-C1832 are 0.21071 and 0.42857 respectively; the adjusted p-value is 0.001786. Using NCBI BLASTn analysis, two sequences were identified in GenBank (accession numbers AY485259 and AY496639) having homology with loci G340 and GP63.

Acknowledgements

This project was supported by National Research Initiative Competitive Grant no. 2003-35206-12834 from the USDA Cooperative State Research, Education, and Extension Service and by a grant from the County of Barnstable Massachusetts, USA (both to SBR).

References

Arnold, WS. 2001 Bivalve enhancement and restoration strategies in Florida, U.S.A. Hydrobiologia, 465(1-3): 7-19

Goudet J (1995) FSTAT version1.2: a computer program to calculate Fstatistics. Journal of Heredity, 86: 485-486. Version 2.9.3 available from http://www.unil.ch/izea/softwares/fstat.html.

Kantety RV, La Rota M, Matthews DE, Sorrells ME. (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology. 48(5-6):501-510

National Marine Fisheries Service. 2004. Annual Commercial Landing Statistics. [Online]. http://www.st.nmfs.gov/st1/commercial/landings/annual_landings.html

92 Roberts SB. and Goetz, F.W. 2003. Expressed sequence tags analysis of genes expressed
93 in the bay scallop, *Argopecten irradians*. (Short Report) Biological Bulletin, 205: 227-
94 228.
95
96 Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001)
97 Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L):
98 frequency, length variation, transposon associations, and genetic marker potential.
99 Genome Research, 11: 1441-1452.
100
101 van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker:
102 software for identifying and correcting genotyping errors in mirosatellite data. Molecular
103 Ecology Notes, 4: 535-538.
104
105 Walsh PS, Metzger DA, and Higuchi R. 1991. Chelex 100 as a medium for simple
106 extraction of DNA for PCR-based typing from forensic material. Biotechniques, 10: 506-
107 513.

Locus (Accession #)	Primers	T_{an}	SSR	Size (bp)	N_a	N_i	H_o	H_e	P_{HW}
M26 (CV660848)	F: CACTTTCAGCAGATATTCTTGAGG R: TCCCATCCTCTCCTTCACAG	55	$(GAT)_{10}$	120	5	44	0.591	0.656	0.1938
GL23 (CV828452)	F: ATAAAACAGGCAAAGAGGCAC R: TGCTTGGTGAATGGGGC	55	$(CA)_{11}$	132	5	20	0.25	0.649	0.0063
S336 (CN783139)	F: GCGGAGGCAGATTCTTTCTTTTC R: GGTCGTGGATTGTAAGCATTGTC	54	$(CAG)_5$	128	4	23	0.478	0.608	0.1375
G340 (CN783297*)	F: CGCTTGTGTTTTACGAGGAGAAGG R: TGACGGGGTGTGATGTCTGACC	53	$(GAT)_5$	114	5	47	0.575	0.642	0.1625
GP63 (CK484125*)	F: AACTTTTCCCTCATCGTGTCACC R: CAGTCACAACTATCAACCTGCCC	54	$(CAG)_5$	223	4	20	0.600	0.522	0.2250
N391 (CN782436)	F: TCATCGCCTCCACCTTCAG R: GATCACACTTTGATTTGTCCTACG	58	$(AG)_{14}A(AG)_5$	243	4	22	0.591	0.720	0.0938
C1831 (CK484157)	F: CGAGTATCAATAGCCGAATCTAAGC R: CCGTAGTTAGATCTCTGTTGGTAG	52	$(GGC)_5$	122	2	21	0.095	0.093	0.9625
C1832 (CK484157)	F: CAGTTATGGATCAGGCGGTAGAAG R: GCGAGCGAGTACAACCTTAAAACAC	55	$(GTG)_6$	122	2	23	0.217	0.198	0.7813

Table 1: EST-SSRs in the bay scallop (*Argopecten irradians*)

T_{an}, annealing temperature; N_a, number of alleles; N_i, number of individuals assayed; H_o, observed heterozygosity; H_e, unbiased expected heterozygosity; P_{HW}, probability that genotype proportions conform to Hardy-Weinberg equilibrium. The adjusted value for significance (5%) following Bonferroni corrections is 0.00625. Asterisk indicates that multiple bay scallop ESTs contain given loci. The accession number for the sequences each loci was originally characterized is listed.