Vol. 262: 299–303, 2003 | MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser | Published November 7

NOTE

Isolation and characterization of DNA microsatellites for capelin *Mallotus villosus* population genetic studies

K. H. Røed[1,*], S.-E. Fevolden[2], J. S. Christiansen[2], L. Midthjell[1]

[1]The Norwegian School of Veterinary Science, PO Box 8146, Dep, 0033 Oslo, Norway
[2]Norwegian College of Fishery Science, University of Tromsø, 9037 Tromsø, Norway

ABSTRACT: This study reports on the isolation and characterization of capelin *Mallotus villosus* DNA microsatellites and their potential use in population genetic studies. High numbers of dinucleotide repeats were detected among the isolated microsatellites. Primers and polymerase chain reaction conditions were developed for 17 polymorphic microsatellite markers, and in a test panel of 60 capelin the number of alleles at each locus ranged from 4 to 63, and observed heterozygosity ranged from 0.11 to 0.95. Six loci showed an excess of homozygotes, presumably due to the presence of null alleles. The 11 other loci showed statistically significant population differentiation between samples from the NE and NW Atlantic, while little differentiation was detected within each of these regions.

KEY WORDS: Capelin · *Mallotus villosus* · Microsatellite · Population structure

INTRODUCTION

Capelin *Mallotus villosus* (Müller, 1776) (Osmeridae) is a small (20 cm), pelagic schooling fish native to the marine waters of the Northern Hemisphere. The capelin exhibits a mosaic of spatially or temporally separate spawning aggregation throughout its distribution area. Large shoals of capelin migrate toward the coasts to spawn, and differences in arrival time for spawners, in location of spawning sites, and in depth of spawning, i.e. bank and beach spawners, have led to the hypothesis that there is a genetic component to the differences in life history traits (e.g. Stergiou 1989, Dodson et al. 1991, Carscadden et al. 1997). Studies of the degree of genetic differentiation among populations of capelin have mainly used polymorphic proteins and mtDNA as genetic markers. Whereas large-scale differentiation has been reported, e.g. separating western and eastern Atlantic stocks (Dodson et al. 1991, Birt et al. 1995), investigations on smaller geographic scales have led to less conclusive results (Mork & Friis-Sørensen 1983, Sørensen & Simonsen 1988, Dodson et al. 1991, Roby et al. 1991). The use of microsatellites as genomic DNA markers to analyze the processes of genetic drift and gene flow has increased rapidly over the last years for various species including fish (cf. Carvalho 1998). As yet, however, there are no published reports of microsatellite markers in capelin. Thus, our primary objective for this study was to isolate and characterize microsatellites for capelin population genetic studies.

MATERIALS AND METHODS

Capelin was collected from 3 sites in the NE Atlantic (NEA) and 2 sites in the NW Atlantic (NWA). The NEA samples were caught in the Barents Sea, north of Svalbard (82° N, 8° E) in February 1999 (N = 44), and in 2 fjords in northern Norway: Porsangerfjord, sampled in

*Email: knut.roed@veths.no

September 1997 (N = 97) and 1998 (N = 20); and Balsfjord, sampled in September 1997 (N = 27) and 2000 (N = 41). The NWA samples were from 2 fjords in eastern Newfoundland sampled in July 2000 at Middle Cove in Conception Bay (N = 31) and at Bellevue in Trinity Bay (N = 41).

Approximately 100 to 200 mg of frozen or EtOH-preserved gill tissue were used to extract DNA by the salt-extraction procedure in accordance with Fevolden & Pogson (1997). Microsatellites were isolated following a standard screening protocol. Briefly, capelin genomic DNA was partially digested with *Sau3*AI and size-selected fragments (400 to 800 bp) were isolated and ligated into the *BamH1* site of the BluescriptSK+ plasmid. The library was screened for inserts containing microsatellites with a synthetic $(GT)_{10}$ oligonucleotide end-labelled with γ^{32}PATP. Positive clones in a secondary screening were randomly picked and sequenced with dye terminator chemistry on an ABI Prism 310 automated sequencer. Readable sequences were obtained from 50 clones. Only dinucleotide repeats of 6 or more perfect units were used for further analysis.

Primer pairs flanking dinucleotide repeats were identified in 19 different microsatellites so that the polymerase chain reaction (PCR) products were approximately 100 to 250 bp in length. The forward primers were end-labelled with fluorescence and the PCR were performed in 10 µl reaction mixtures containing 20 to 40 ng of genomic template DNA, 2 pmol of each primer, 50 mM KCl, 1.5 mM MgCl, 10 mM Tris-HCl, 0.2 mM dNTP and 0.5 U AmpliTaq (ABI). A denaturation phase at 94°C for 5 min was followed by 30 cycles of: 95°C for 1 min, 30 s annealing temperature as given in Table 1, and 72°C for 1 min. A final extension of 72°C for 10 min was performed. The PCR products were electrophorised using an ABI Prism 310 automated sequencer. Amplification products were tested for variability in a panel of 60 capelins randomly picked from the 3 sampled locations in the NEA. Sex was determined for 25 individuals (10 males and 15 females).

The computer package GENEPOP, Version 3.1 (Raymond & Rousset 1995), was used to test for genotypic disequilibrium among the loci used and for conformity to Hardy-Weinberg equilibrium (HWE) and for exact *G*-tests of allele frequency homogeneity both among sampling years within populations and among populations for pooled values over years. All tests used Markov chain algorithm as implemented in the program (dememorization 1000, batches 100, iterations per batch 1000). FSTAT, Version 2.9.3.2 (Goudet 1996), was used for estimates of allele richness and Nei's (1987) unbiased estimates of heterozygosity within each population, and for tests of differences in these parameters among populations in the NEA and NWA. The extent of population differentiation was quantified by Weir & Cockerham's (1984) estimator of pair-wise F_{ST} values using bootstrap and jack-knife resampling procedures and Bonferroni corrections for pair-wise significance, as implemented in this program.

RESULTS

Microsatellite repeats were relatively common in the capelin genome. Sequencing of 50 randomly picked positive clones yielded 42 with perfect dinucleotide repeats with 6 or more tandemly repeated units. Among them, 16 contained 2 or more clearly separated dinucleotide repeats. Several clones contained large microsatellites, the longest consisting of 100 repeat units (Table 1). For some clones we were unable to sequence through the insert due to long repeat arrays. Among the 19 primer sets designed, 2 produced no product in the PCR, while 17 gave satisfactory amplification results and showed polymorphism within the test panel. Table 1 shows repeat sequences, primer sequences, annealing temperature, length of fragments amplified, number of alleles and observed and expected heterozygosities for the 17 microsatellites obtained for the test panel. All 60 individuals could be scored in all microsatellites except Mav-93, for which only 21 individuals amplified a product. No loci appeared to be duplicated (i.e. presence of more than 2 alleles within an individual).

The 17 microsatellite loci showed a wide range of genetic variability, with the number of alleles detected at each locus ranging from 4 to 63 and observed heterozygosity ranging from 0.11 to 0.95 (Table 1). The mean values of number of alleles and expected heterozygosity were 20.4 and 0.84 respectively. Six loci showed highly significant ($p < 0.0001$) deviation from HWE (Table 1). An excess of homozygous individuals was detected in these loci, probably indicating presence of null alleles due either to no amplifiable alleles or to drop out of larger alleles owing to competitive PCR amplification of smaller alleles. The deviation in Mav-4 and Mav-93 may possibly be due to sex-linked location of these loci as no heterozygous males were detected. Heterozygous males were recorded for the other microsatellites, suggesting autosomal inheritance for these loci. The 6 loci that highly deviated from HWE were not used in further population differentiation analyses.

The exact tests for sample differentiation across the 11 loci used revealed no significant differences in allele frequency distribution between sampling years in Porsangerfjord ($p = 0.45$) and Balsfjord ($p = 0.47$). The data for different years within those fjords were therefore pooled for subsequent analyses. All 11 loci

Table 1. *Mallotus villosus*. Characterization of 17 capelin microsatellite loci tested on 60 randomly picked individuals from the 3 sites in the NE Atlantic. GeneBank accession numbers are AY291349–AY291365. T_a = optimized annealing temperature; N_{ALL} = number of alleles detected; H_o = observed heterozygosity; H_e = expected heterozygosity. H_o is significantly different from Hardy-Weinberg equilibrium (HWE) at $^{*}p < 0.05$ and $^{***}p < 0.0001$

Locus	Repeat sequence	Primer sequences 5′ → 3′	T_a (°C)	N_{ALL}	Allele size range (bp)	H_o	H_e
Mav-2	$(CA)_{19}$	F: TAG GCA TGT CAG AAC TGG AC R: CAT GTG TCA TCG CCA AAA	45	17	101–143	0.58***	0.90
Mav-4	$(GT)_{11}$	F: GTG CGT GCC AGT AGA AA R: TCA GTC ACA CTT ACC GAG AG	45	15	99–132	0.43***	0.81
Mav-9	$(CA)_{30}$	F: ATG TGT GAG GCC AGA GCA GT R: CCC ACA CCT GAG ACA GAC C	50	29	168–258	0.95	0.96
Mav-15	$(CA)_7AG(CA)_{11}$	F: CAA CAG CCT TTT TAT ACA CA R: TAA GGG ATT GTT TAG AAT GA	45	13	159–193	0.68*	0.85
Mav-17	$(GT)_{10}$	F: GGG CAA AGC ATT GTC TGA R: ATC ATT CCT GAG GGC TAC AG	55	14	186–236	0.78	0.76
Mav-20	$(CA)_9GATA(CA)_6$	F: GCC GAC CAA CTC TTC TCT R: TTC ACT GAC TCA CTG GAA AC	50	26	121–188	0.56***	0.95
Mav-38	$(CA)_{15}$	F: GCA ACA TGA CAG GAC TCG TT R:GGG CAA GGC TAA AGA AGA AA	45	16	131–173	0.80	0.86
Mav-42	$(GT)_{13}(TA)_3$	F: GCA TAG TGT CCT GAA TGA TG R : GTG ACA CTT TGC TTG GAG	45	10	103–123	0.64	0.63
Mav-51	$(CA)_7TGTC(CA)_{12}$	F: TGA TAG TCT GGG AGG TTT GG R: GGG TGG GCG TCG TTT	50	12	170–206	0.63	0.74
Mav-53	$(CA)_{100}$	F: AGT ACC AGT TTT GCC CAA CA R: CCA GGC CTC ATC TCT TTC A	45	63	85–348	0.55***	0.99
Mav-56	$(CA)_{16}$	F: CCC CAG CCT CCC TCA GTA R: CCT TTA CCC AGA GTG ACC AT	45	26	89–145	0.93	0.96
Mav-62	$(CA)_{19}$	F: CAA GTG TGC TGG GAT GAA GA R: ACG CTG CAG GAG TCC AAC	50	19	109–147	0.93	0.91
Mav-81	$(GT)_8$	F: ATG TGA GAC ACG CAT ACA CT R: TGT GCA AAC CAG AAT GAA T	50	4	118–130	0.38	0.38
Mav-93	$(GT)_{11}$	F: GGT TAA CCT GAG ATA ATA CA R: AAA AAG GAT AAA GGA GAC	45	9	102–126	0.11***	0.78
Mav-99	$(CA)_{17}$	F: GGC TGC CAT TCC TCC ACC TG R: CGG GGG CCA GCT TTC AGA CT	50	21	60–122	0.95	0.94
Mav-123	$(GA)_{38}$	F: GCG TCT GTG TGC GGA TTG R: GTG TGA CCC AGC CCT ACT GC	50	39	147–264	0.83***	0.97
Mav-135	$(CA)_{21}$	F: GAC GTC ACT CCC GCA AGG R: CAG CGT TAC CAG GCA AAT CC	50	13	89–121	0.88	0.82

were polymorphic with high allelic diversity and heterozygosity in all samples. Altogether 274 different alleles were recorded, with mean number of alleles over loci ranging from 14 in Conception Bay to 20 in Porsangerfjord. The mean of Nei's (1987) unbiased estimates of heterozygosity ranged from 75% in Trinity Bay to 81% in Barents Sea. Estimates of allelic richness, which is a measure of the number of alleles independent of sample size, gave very similar values for the 5 sampling sites, ranging from 11.4 in Balsfjord to 11.7 in Conception Bay. Tests of differences in gene diversities among the NEA samples and the NWA samples revealed no significance in estimates of allele richness, observed heterozygosity or Nei's (1987) unbiased estimates of heterozygosity.

With 5 samples and 11 loci, there were 55 tests for HWE among which 5 were significant ($0.01 > p > 0.05$), which is similar to that expected due to Type 1 error. The 5 significant deviations were all spread among different loci. The global tests across all loci revealed no significant deviation for any of the sampled populations. Exact tests for linkage disequilibrium between microsatellite loci across all populations showed no significant linkage. The allele frequency differences among the sampled populations were not large, yet the exact tests for population differentiation revealed significant differences ($p < 0.001$) for all loci combined and for 10 out of the 11 loci independently, most of which were between the NEA and the NWA sites (Table 2). F_{ST} showed statistically significant popula-

Table 2. *Mallotus villosus*. Genetic differentiation among samples of capelin from 3 locations in the NE Atlantic (NEA; Barents Sea, Porsangerfjord, and Balsfjord), and from 2 locations in the NW Atlantic (NWA; Conception Bay and Trinity Bay). The estimator of pair-wise F_{ST}-values are given below the diagonal, and the significance levels (p) of the exact test for genetic differentiation for each population pair across all loci are given above the diagonal. Number of loci with significant ($p < 0.05$) differences is given in parentheses

	NEA			NWA	
	Barents Sea	Porsangerfjord	Balsfjord	Conception Bay	Trinity Bay
Barents Sea		0.013 (1)	0.064 (2)	<0.001 (7)	<0.001 (8)
Porsangerfjord	0.0014		0.005 (4)	<0.001 (11)	<0.001 (11)
Balsfjord	−0.0004	0.0019		<0.001 (11)	<0.001 (11)
Conception Bay	0.0442	0.0355	0.0324		0.062 (2)
Trinity Bay	0.0345	0.0284	0.0270	0.0006	

tion differentiation for all pair-wise comparisons between samples from the NEA and NWA with a mean (±SE) F_{ST} value of 0.030 (±0.010).

The exact tests for population differentiation over all loci within the NEA and NWA gave small, yet statistical significant, differences between Barents Sea and Porsangerfjord ($p = 0.013$) and between Porsangerfjord and Balsfjord ($p = 0.005$). Both the exact tests and the F_{ST} estimates showed that within the NEA the differences between the 2 Norwegian fjords were generally greater than when comparing the fjords and the Barents Sea (Table 2). Four loci showed significant differences between the 2 fjords while only 1 locus showed significant difference between the Barents Sea and the Porsangerfjord. Neither the exact tests over all loci nor the F_{ST} statistics revealed significant differences between the samples from the Barents Sea and Balsfjord or between the 2 samples in the NWA.

DISCUSSION

In this study we report the isolation and development of original microsatellite loci for use in capelin population discrimination. Many of the capelin microsatellites were characterized as being relatively long and with high allele variation. This supports the trend that fish has longer microsatellites and more alleles than other vertebrate classes (DeWoody & Avise 2000, Neff & Gross 2001). In the present reference panel of capelin, the mean number of alleles and heterozygosity were 20.4 and 0.84 respectively, similar to values reported for marine fishes by DeWoody & Avise (2000) (mean number of alleles = 19.9 and mean heterozygosity = 0.77). These authors reported that marine fish species have greater microsatellite allele variation than freshwater and anadromous fish species and that this is consistent with the increased evolutionary effective population sizes of marine fishes. The effective population size of capelin is in the order of billions for large oceanic stocks (Gjøsæter 1998).

Eleven of the 17 loci that conformed to Hardy-Weinberg expectations showed no evidence of linkages and covered a broad range of degree of polymorphism, which suggests that these loci are very suitable for studies of population divergence in capelin. Possibly null allele and/or 'drop out' artefacts reduced the reliability of population analysis for the other 6 loci. The allele frequency distribution in the 11 remaining microsatellite loci is indicative of relatively little genetic differentiation among the sampled sites within both the NEA and the NWA. Highly significant differentiation was detected between these 2 regions. Significant genetic divergences of capelin between the NEA and NWA together with minor differentiation within each of these regions are in accordance with previous reports (Mork & Friis-Sørensen 1983, Dodson et al. 1991, Roby et al. 1991, Birt et al. 1995). To determine whether the low level of differentiation detected has biological significance, we have refined our sampling strategy by expanding the geographic area on a large scale and by including more fjords with presumably local stocks of capelin.

Acknowledgements. J. Carscadden, Department of Fisheries and Oceans (DFO), St John's, Canada, is thanked for providing samples from Newfoundland. The study was supported by grant nos. 127546/120 and 140290/140 (BASECOEX) from The Research Council of Norway.

LITERATURE CITED

Birt TP, Friesen VL, Birt RD, Green M, Davidson W (1995) Mitochondrial DNA variation on Atlantic capelin, *Mallotus villosus*: a comparison of restriction and sequence analyses. Mol Ecol 4:771–776

Carscadden JE, Nakashima BS, Frank KT (1997) Effects of fish length and temperature on the timing of peak spawning in capelin (*Mallotus villosus*). Can J Fish Aquat Sci 54: 781–787

Carvalho GR (1998) Advances in the molecular analysis of fish population structure. Ital J Zool 65:21–33

DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. J Fish Biol 56:461–473

Dodson JJ, Carscadden JE, Bernatchez L, Colombani F (1991) Relationship between spawning mode and phylogeographic structure in mitochondrial-DNA of North-Atlantic capelin *Mallotus villosus*. Mar Ecol Prog Ser 76:103–113

Fevolden SE, Pogson GH (1997) Genetic divergence at the synaptophysin (*Syp*-1) locus among Norwegian coastal and north-east arctic populations of Atlantic cod, *Gadus morhua*. J Fish Biol 51:895–908

Gjøsæter H (1998) The population biology and exploitation of capelin (*Mallotus villosus*) in the Barents Sea. Sarsia 83: 453–496

Goudet J (1996) FSTAT (Version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486

Mork J, Friis-Sørensen E (1983) Genetic variation in capelin *Mallotus villosus* from Norwegian waters. Mar Ecol Prog Ser 12:199–205

Neff BD, Gross MR (2001) Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. Evolution 55:1717–1733

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Roby D, Lambert JD, Sévigny JM (1991) Morphometric and electrophoretic approaches to discrimination of capelin (*Mallotus villosus*) populations in the estuary and Gulf of St. Lawrence. Can J Fish Aquat Sci 48:2040–2050

Sørensen EF, Simonsen V (1988) Genetic differentiation among populations of capelin *Mallotus villosus* from the west coast of Greenland. J Appl Ichthyol 4:23–28

Stergiou KI (1989) Capelin *Mallotus villosus* (Pisces: Osmeridae), glaciations, and speciation: a nomothetic approach to fisheries ecology and reproductive biology. Mar Ecol Prog Ser 56:211–224

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38: 1358–1370

Editorial responsibility: Otto Kinne (Editor), Oldendorf/Luhe, Germany

Submitted: September 26, 2002; Accepted: September 9, 2003
Proofs received from author(s): October 17, 2003