Conservation Genet Resour (2012) 4:71–74
DOI 10.1007/s12686-011-9477-5

TECHNICAL NOTE

Development of 61 new transcriptome-derived microsatellites for the Atlantic herring (*Clupea harengus*)

A. G. F. Teacher · K. Kähkönen · J. Merilä

Received: 15 June 2011 / Accepted: 26 June 2011 / Published online: 13 September 2011

Abstract Next generation sequencing has made it possible to develop large numbers of molecular markers for non-model organisms quickly and at relatively low cost. We identified and tested 155 potential new transcriptome-derived microsatellite markers for the Atlantic herring (*Clupea harengus*) using 454-sequencing. By testing the markers in five individuals from each of eight populations, we found that 61 of these markers are potentially useful. Further tests in 25 individuals from each of three populations indicate that 50 of these markers are polymorphic, highly reliable for scoring and show no indication of consistent heterozygote deficiency or linkage disequilibrium. Transcriptome-based microsatellite markers may prove particularly useful for studies of adaptation and population differentiation in species such as herring which live in highly variable habitats.

Keywords Microsatellite · 454 · Herring · *Clupea harengus* · *Clupea pallasii* · Fish · Transcriptome

Microsatellite markers are used extensively within molecular ecology and conservation genetics, as well as in stock assignment and assessments of genetic diversity for commercial fish (e.g. Hansen et al. 2001a, b; Perez-Enriquez et al. 1999). One of the most commercially important species in the North Sea and Baltic Sea is the Atlantic herring (*Clupea harengus*; ICES 2011). However, remarkably few genetic resources have been developed for this species—there are currently nine Atlantic herring (McPherson et al. 2001), and 14 Pacific herring *Clupea pallasii* (Olsen et al. 2002) microsatellite markers. In this study, we used next generation sequencing to develop transcriptome-derived microsatellite markers for the Atlantic herring.

Two herring (one of German and one of Finnish origin) were obtained from a public aquarium (Sea Life Helsinki), and euthanized with MS222. Total RNA from 30 mg of gill tissue was extracted using the RNeasy Minikit (Qiagen, Finland); the resulting concentrations were 371.3 and 332.6 ng/µl. cDNA was synthesized using the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen). The concentrations were equalised, the samples were barcoded and pooled, and sequenced on half a plate using a GS FLX machine (454 Life sciences/Roche) by the University of Helsinki sequencing service. Barcodes and poly-A tails were removed from the data, and the sequences (416,521 reads, mean length 283.7 base pairs) were assembled using MIRA 3.2.1 (Chevreux et al. 2004). This assembly (unpadded, 33,701 contigs, mean length 315.8 bp, max length = 5,284 bp) was used with QDD (Meglécz et al. 2010) for the identification of microsatellites and primers. In total, 266 simple markers and 75 compound markers with primer binding sites and ≥5 repeat units were identified. Of these, 155 simple markers were chosen for testing based on either having (1) ≥6 repeat units of a motif of any size, or (2) five repeat units of motifs between three and six base pairs long.

The 155 markers were first tested on eight individuals, consisting of two individuals from each of four populations (SIM: Finland, 65°37′N, 24°52′E; P: Estonia, 59°03′N, 22°28′E; RUG: Germany, 54°34′N, 13°27′E; SPA: Sweden, 59°5′N, 11°14′E). Total DNA was extracted using a silica-based method (Ivanova et al. 2006). Primers were

A. G. F. Teacher (✉) · K. Kähkönen · J. Merilä
Ecological Genetic Research Unit, Department of Biosciences, University of Helsinki, PO Box 65, Biocenter 3, Viikinkaari 1, 00014 Helsinki, Finland
e-mail: amber.teacher@helsinki.fi

Table 1 Transcriptome-derived microsatellites

Code	Accession no.	n (alleles)	Range	H_E (P)	H_O (P)	H_E (SIM)	H_O (SIM)	H_E (RUG)	H_O (RUG)	Dye	Forward primer (5′–3′)	Reverse primer (5′–3′)	Motif
Her1	JN182998	3	126–134	0.527	0.696	0.518	0.605	0.505	0.696	FAM	TGCCCCTTACAACATCTTTG	AACGATCACAAAACACGGCT	$AGCT_5$
Her8[a]	*JN183054*	*3*	*252–263*	–	–	–	–	–	–	*HEX*	*TCATCAGTGTGCTCGTTTCC*	*CCTGTCCATCGTCCTGTCTT*	$CCTGT_5$
Her10[a]	*JN182999*	*3*	*119–127*	–	–	–	–	–	–	*FAM*	*GCTTATTTGCAGACACGCAC*	*TCGACAAACAACGTATTCACTG*	$TGAC_5$
Her12	JN183011	5	137–151	0.462	0.614	0.541	0.548	0.339	0.424	TET	TTCTTCTGCAGCTGGGATTT	CCTTCCTTCATGTCCTGTCTG	$ACAA_5$
Her14	JN183020	5	114–138	0.474	0.429	0.482	0.488	0.539	0.538	HEX	GATTTCACCTCCTTTTGCCA	CTTGTACGTCGCTTTGGACA	$CATTTA_5$
Her18	JN183025	4	166–179	0.512	0.556	0.519	0.512	0.538	0.674	TET	GCTGACTGAATGGACTGACTGA	CAAGAGCCAGACCAGTGACA	$TGAT_5$
Her20	JN183026	6	92–107	0.335	0.356	0.457	0.444	0.375	0.429	HEX	GGTGTGGCCCAGTCACAA	CTATGTACGGGCTCTGTGGC	ATA_5
Her21	JN183027	4	146–157	0.282	0.319	0.312	0.283	0.217	0.239	TET	TGGCCTTGTGACGAACACTA	TGCTGCTTATGGACTGTTATTCC	ATT_5
Her22	JN183028	2	119–122	0.286	0.341	0.426	0.447	0.344	0.391	FAM	GCAATCCCAGAATGGACTGT	CCGATCATAAATGAACATGCAA	AAT_5
Her24[b]	*JN183029*	*2*	*134–137*	–	–	–	–	–	–	*TET*	*TCCAGTGTAAAATGCATCTAAAAGA*	*TGAGAACTATTGGGATGGGG*	TAA_5
Her25	JN183030	4	145–154	0.422	0.444	0.494	0.378	0.452	0.550	FAM	TGGCACTCTGAGATTCTTCTG	AAACGCACACCTAGGGACAC	ATT_5
Her28	*JN183031*	*11*	*86–124*	*0.763*	*0.289**	*0.756*	*0.302**	*0.795*	*0.429**	*FAM*	*CATTCATCCATCTCATCCCTAAC*	*GAGGAGATATGGAGACTCAGGG*	$CTGT_6$
Her33[a]	*JN183032*	*7*	*123–152*	–	–	–	–	–	–	*TET*	*GCCCAACACCCTTATCTTCA*	*GTCCTCACTTAGCTCACCGC*	$GAGG_6$
Her34[a]	*JN183033*	*5*	*211–216*	–	–	–	–	–	–	*FAM*	*AATCCCACATAACCCCAACA*	*TATGAATTCTCTGCCCTCCC*	TC_6
Her36	JN183034	5	163–179	0.631	0.739	0.639	0.644	0.713	0.707	TET	AAACAGATTCCCCTCGTCCT	AAACGGTTCGGTGCTATTTG	$AATC_6$
Her37	JN183035	4	114–122	0.062	0.064	0.085	0.065*	0.021	0.021	FAM	TTGAGCTCAGGGGAAAATTG	GGGCTGGTTTTTGTTGTACTC	AAG_6
Her38[a]	*JN183036*	*8*	*111–139*	–	–	–	–	–	–	*HEX*	*TGGTGGAGAGTACCCAAAGG*	*TGCTATGCCAAGACAACCAA*	$TATT_6$
Her40	JN183037	3	178–182	0.178	0.196	0.104	0.109	0.065	0.067	FAM	TTGTGCATTCTGCACTAGCC	CCACGTATGAACTTCCTGGG	CA_6
Her41	JN183038	2	115–118	0.299	0.311	0.411	0.391	0.497	0.617	HEX	TTTCAACATGGCTGGCAATA	GAGTTCCGCCCAACAAGTTA	ACA_6
Her43	JN183039	6	171–184	0.103	0.106	0.043	0.043	0.339	0.364	FAM	TTAGCCCTGCAAGCCAATTA	TTTCCCTTCAGACGAACACC	AG_6
Her50	JN183040	6	108–142	0.149	0.154	0.203	0.174	0.341	0.386	HEX	GGTGAGTCCAGGGAGAACAC	ATTGTTTTGTAATTTATTGTACTGCAC	$TAGAA_6$
Her51[a]	*JN183041*	*4*	*243–255*	–	–	–	–	–	–	*TET*	*ACCATACCCAGACAACCACC*	*GCGTTGTTTATACCGGCAGT*	$AAAT_6$
Her53	JN183042	6	114–127	0.558	0.575	0.541	0.512	0.568	0.611	HEX	TCCCTCTCCTACATTCGGTG	TGTTCAGGATCTCACCCCTC	TGT_6
Her54	JN183043	7	166–206	0.700	0.659	0.725	0.674*	0.678	0.698	TET	AACTGGCTGTGGACATGGAT	TCTGTCTGCCAGGGGAATAC	TGG_6
Her58	JN183044	4	94–100	0.183	0.109*	0.235	0.163*	0.246	0.277	FAM	CATGTTCAATACTTATTTTACCCGC	TCCGCATTGTTTATCTGTCAA	TA_6
Her59	JN183045	3	245–249	0.462	0.568	0.507	0.478	0.539	0.639	HEX	ATTACCCAGGGGACACATCA	AAAGGTGCTGATGATGGAGG	CA_6
Her61	JN183046	6	104–129	0.588	0.564	0.693	0.432*	0.532	0.500	FAM	TTTTGCTTCTGGTGTTGCTG	GGTTGGCCTAAGACATCTATTAAA	TGT_6
Her62	JN183047	3	170–174	0.258	0.295	0.161	0.174	0.305	0.370	HEX	CCCTGAAAAGACTCTGGCTG	GCACAGATTAGTGCCACCCT	CA_6
Her63	JN183048	2	102–104	0.082	0.085	0.169	0.182	0.160	0.170	TET	CCTGAACAGTTTTGTATGAGGC	GGAGAAAACTCACTGTGGGC	TC_6
Her64	JN183049	4	97–109	0.608	0.630	0.565	0.641	0.558	0.564	FAM	AGAGCCAGACTCTCAGCAGG	TTAAAAGAGTGAAGGGGCCA	$AAAC_6$
Her67	JN183050	5	152–165	0.375	0.370	0.294	0.289	0.338	0.386	FAM	CGTGCTCAAAGGTAGAACCG	CGATCTTCGTTTTGGTCCTC	TAT_6
Her71	JN183051	3	85–97	0.484	0.512	0.484	0.467	0.516	0.511	HEX	ACATCACCCGACTGCTAACC	GTGGCTCTGGATGATGGTCT	$AGGC_6$
Her73	JN183052	4	141–153	0.121	0.128	0.165	0.130	0.160	0.170	FAM	GGAAGGCTGACGGTGAAGTA	AAAGGTCCTCAGAAGCCTCC	GCT_6
Her77	JN183053	8	164–181	0.343	0.386	0.519	0.545	0.542	0.595	HEX	ACAGGGCGATTAATGGACAA	AAATGATCCTCCCAAAAGGG	$TTCA_7$
Her84	JN183055	2	149–151	0.142	0.149	0.123	0.130	0.353	0.391	TET	CAGGTATGGGGTGTGCTTCT	AGAGGCGTCTTTGAGAGTGC	CA_7
Her89[a]	*JN183056*	*14*	*197–244*	–	–	–	–	–	–	*HEX*	*GCACTTCTCCTGCTCTTGCT*	*CACTGGCTGATGTTTGTGCT*	CA_7
Her97	JN183057	6	88–102	0.627	0.630	0.576	0.511	0.605	0.564	FAM	GCTGGGTTTATTCTTGTATGGG	GCTACATATCAAACACACACATCA	TG_7
Her98	JN183058	3	135–139	0.144	0.152	0.219	0.200	0.112	0.116	HEX	ACAATCAGCATGACGGATGA	CACAGCATTCGTCTGAAAGG	CA_7
Her100	JN183000	9	123–141	0.863	0.721	0.738	0.652	0.810	0.794	FAM	ACGTTCATTCTTGCCAAACC	ATGCCAAATTCCCAATGAAA	TG_7
Her101	JN183001	2	89–93	0.143	0.149	0.144	0.109	0.267	0.304	HEX	TGGTGTTGGTTAATGTAATGCC	TGCCTAGAGCTTCATCCAGTT	AC_7

Table 1 continued

Code	Accession no.	n (alleles)	Range	H_E (P)	H_O (P)	H_E (SIM)	H_O (SIM)	H_E (RUG)	H_O (RUG)	Dye	Forward primer (5′–3′)	Reverse primer (5′–3′)	Motif
Her102	JN183002	5	94–108	0.171	0.182	0.186	0.200	0.228	0.250	TET	GGCCTTGAGAAAGCACTGAA	TGATTTGCTGGCATGTGTTT	AC_7
Her104	JN183003	14	146–198	0.895	0.800	0.871	0.733	0.883	0.813*	HEX	CCGAGGTAAAGTTCACTGCAA	AAGCCACGGATGAGGAAAC	TC_8
Her107	JN183004	7	87–127	0.167	0.130*	0.268	0.289	0.132	0.136	HEX	TCAGACTTCAGGAAATGACAGAA	GGATGCCTCATTCTCTGGTG	AG_8
Her109	JN183005	7	123–141	0.849	0.739*	0.868	0.818	0.762	0.771	FAM	TTCAGCGATTAGTCAGTACAGCA	TGGATCACTGTACATGTCATTTCTT	TAA_8
Her111	JN183006	4	123–135	0.515	0.474	0.540	0.600	0.465	0.442	TET	CCTGTCTGTGCTGTGAAGGA	TGTCTTAATTGTGGCGCATC	AG_8
Her114	JN183007	7	164–183	0.609	0.537	0.644	0.565	0.636	0.737	TET	CCCACTCACAAAGGCATCTT	TGATATTGGCTCGGGACTTC	TTG_8
Her117	JN183008	3	113–117	0.161	0.174	0.022	0.022	0.144	0.152	TET	TGCGTGAGTTGGTGAGAAGA	CCTCATAGAAACCGCAAAGC	GA_8
Her118	JN183009	5	135–151	0.388	0.422	0.357	0.298	0.243	0.267	FAM	CATGTGGTCTAGCTTTTCTTTCC	TGTTCTGATTTCTTGCCCTG	CT_9
Her119	JN183010	7	92–115	0.781	0.674	0.773	0.674	0.696	0.773	HEX	TGATTGCATCTCTGTGTTTGTT	ACTGGAAAGGCGTGAATGTC	TG_9
Her121[b]	*JN183012*	*2*	*144–146*	–	–	–	–	–	–	*FAM*	*TGCCTCCTAGAAGCAGACAAA*	*CATCACAAGACAAATTACATTACCA*	*TC_9*
Her124	JN183013	3	98–110	0.186	0.200	0.324	0.304	0.251	0.283	FAM	TCACTGCATCCTGAGTTCCA	TGAAGGTCATAGAACTTGTGTAATTG	TC_9
Her126	JN183014	4	218–226	0.502	0.628	0.569	0.543	0.580	0.545	TET	GTGACCATCCAGGAAGCAAT	GGGCACCTCATGTACACCAT	TG_9
Her130	JN183015	12	105–140	0.626	0.674	0.737	0.652	0.752	0.919	FAM	CACGGCCTGTTTGTTTTCTT	CGCACCATGTGTCCAAACTA	TG_{10}
Her132	JN183016	3	140–144	0.490	0.478	0.486	0.533	0.482	0.652	TET	AGCGCTGCTTTGACTGTCTC	CGAAACCACAGAGGGTTGTT	TG_{10}
Her133	JN183017	11	87–117	0.625	0.609	0.677	0.711	0.576	0.614	FAM	CCACACTGACTACTATGTGAACAGC	CTTTATATGGGGTATAACAAACAAACG	TG_{10}
Her136	JN183018	3	88–94	0.296	0.340	0.410	0.413	0.281	0.298	FAM	ATCAAACAACCCAGGTCAGC	TGTAAGTCTTCGCCACATGC	AG_{10}
Her139[b]	*JN183019*	*3*	*88–92*	–	–	–	–	–	–	*FAM*	*ACGTTTGCATGTGACTGAGG*	*TGTTTATTGAGTTTCTCAGTGCTGT*	*CA_{11}*
Her140	JN183021	8	102–138	0.819	0.773	0.809	0.870	0.824	0.854	HEX	TTATGTAATGCACATTCAGATTATTTT	TGTCCATCTCTATCTATCTGCCG	$GATA_{11}$
Her141	JN183022	9	186–206	0.813	0.833	0.834	0.825	0.786	0.884	TET	AGGAAAAACTCCCCTGCTGT	TCCTCAGTTTATTTGTGCTCTCA	CT_{11}
Her142	JN183023	10	85–108	0.771	0.674	0.742	0.717	0.612	0.702	FAM	TATGTGAGAGGAAAGCCGGT	TTCAGGAAAGATGGTGGAGG	TG_{12}
Her143	JN183024	10	85–110	0.839	0.756	0.808	0.738	0.870	0.951	HEX	ACATGTGCATAGACGTACAAGTTT	CGTGTGGACAAATAGAAGCG	TG_{12}

The table shows the locus name code, accession number, number of alleles, allele size range, observed heterozygosity (H_O), expected heterozygosity (H_E), dye used for genotyping, primer sequences (5′–3′), and repeat motif. The number of alleles and range are based on results from eight individuals from each of five populations, whilst H_O and H_E are based on results from 25 individuals from each of three populations: P, SIM, RUG. Out of the total of 61 markers, 50 were highly reliable with good levels of polymorphism, easy scoring, no linkage disequilibrium, and no consistent problems with heterozygote deficiency; the markers with potential problems are in italics

[a] Problematic to score

[b] Very low levels of polymorphism in the tests with eight individuals from five populations

* Significant ($P < 0.05$) heterozygote deficiency

labelled with FAM, HEX, or TET (DNA Technology A/S, Denmark), and 10 µl PCRs were performed using 2 pmol of each primer, 1× Phusion® Flash High-Fidelity PCR Master Mix (Finnzymes) and ~2 ng of DNA. The cycling profile was: 98°C for 1 min, followed by 34 cycles of 98°C for 1 s, 58°C for 12 s, 72°C for 20 s, and a final extension at 72°C for 1 min. The PCR products were analysed using a MegaBace 1,000 capillary sequencer and Fragment Profiler 1.2 software (GE Healthcare, Life Sciences). Markers that did not amplify, could not be scored, or did not show ≥2 alleles were removed from the marker set (94 loci removed). The reduced marker set (61 loci, Table 1) was then amplified in 40 individuals, comprising of eight individuals from five populations (four populations as above and FVI: Finland 60°30′N, 27°45′E). Three of these markers had very low polymorphism (i.e. close to fixation of a single allele across populations), and three other markers showed problems with scoring (i.e. single base pair changes or non-specific binding). These six markers with problems were discarded, and the remaining 55 markers were tested with 25 individuals from each of three populations (P, RUG, SIM), to test for Hardy–Weinberg Equilibrium (HWE) and linkage disequilibrium (LD) using GenePop on the Web (Raymond and Rousset 1995). Four of these markers were not possible to score reliably and so were removed at this stage, leaving 51 markers. Tests for HWE identified eight loci with heterozygote deficiency in at least one population, however only one locus showed consistent problems across all three populations (Her28, Table 1). Tests for LD showed no locus pairs with significant LD in more than one population. The final count for highly reliable markers was 50.

As the newly developed markers are within transcribed genes, they may not follow neutral expectations, though few deviations were observed in this study. However, because of their linkage to transcribed genes, they may be particularly useful for looking at population differentiation related to local adaptation. This application is potentially important for identifying population units for fisheries management, and later for stock assignments. Comparisons between similar studies (e.g. Csencsics et al. 2010; Mikheyev et al. 2010; Nair et al. 2011) are difficult because genomes of different taxa vary substantially in size and microsatellite composition, and also because many publications do not specify their criteria for determining a potential marker. However, it is clear that the technique of mining small 454 sequencing runs for microsatellites is highly effective.

Acknowledgments Thanks to Erkki Jokikokko, Anti Vasemägi, Mike Luner, Carl André, Eric Hoffmann, Anna-Karin Ring, Timo Jääskeläinen, Teija Aho and Jukka Pönni for help in providing the samples. Thanks also to Marika Karjalainen for lab work for the 454 sequencing run, and to Sea Life Helsinki for providing the herring for the initial cDNA preparation. Fish used in this study were not subjected to any experimental procedures, and were dispatched using approved methods according to Annex 3 of the Decree of the Ministry of Agriculture and Forestry (Finland). This project was funded by the Academy of Finland (grant no. 138043 to AGFT, and grant nos. 129662 and 134728 to JM) and from the European Community's Seventh Framework Programme (FP/2007-2013) under grant agreement n° 217246 made with the joint Baltic Sea research and development programme BONUS (grant to JM).

References

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14:1147–1159

Csencsics D, Brodbeck S, Holderegger R (2010) Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. J Hered 101:789–793

Hansen MM, Kenchington E, Nielsen EE (2001a) Assigning individual fish to populations using microsatellite DNA markers. Fish Fish 2:93–112

Hansen MM, Ruzzante DE, Nielsen EE, Mensberg KLD (2001b) Brown trout (*Salmo trutta*) stocking impact assessment using microsatellite DNA markers. Ecol Appl 11:148–160

ICES—International Council for the Exploration of the Sea (2011) http://www.ices.dk/marineworld/fishmap/ices/default.asp?id=Herring. Accessed 03 Feb 2011

Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. Mol Ecol Notes 6:998–1002

McPherson AA, O'Reilly PT, McParland TL, Jones MW, Bentzen P (2001) Isolation of nine novel tetranucleotide microsatellites in Atlantic herring (*Clupea harengus*). Mol Ecol Notes 1:1–32

Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26:403–404

Mikheyev AS, Vo T, Wee B, Singer MC, Parmesan C (2010) Rapid microsatellite isolation from a butterfly by De Novo transcriptome sequencing: performance and a comparison with AFLP-derived distances. PLoS One 5:e11212

Nair A, Kumar KS, George S, Gopalan SV, Li M-H, Leder EH, Merilä J (2011) Sixty-two new microsatellite markers for an endemic frog *Indirana beddomii* from the Western Ghats biodiversity hotspot. Conserv Genet Res 3:167–171

Olsen JB, Lewis CJ, Kretschmer EJ, Wilson SL, Seeb JE (2002) Characterization of 14 tetranucleotide microsatellite loci derived from Pacific herring. Mol Ecol Notes 2:101–103

Perez-Enriquez R, Takagi M, Taniguchi N (1999) Genetic variability and pedigree tracing of a hatchery-reared stock of red sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. Aquaculture 173:413–423

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