

## PRIMER NOTE

# Seventy-five EST-linked Atlantic salmon (*Salmo salar* L.) microsatellite markers and their cross-amplification in five salmonid species

ANTI VASEMÄGI,\*† JAN NILSSON\* and CRAIG R. PRIMMER‡

\*Department of Aquaculture, Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden, †Institute of Zoology and Hydrobiology, University of Tartu, 51014 Tartu, Estonia, ‡Department of Biological and Environmental Sciences (Biocentre 3), PO Box 65, FIN-00014, University of Helsinki, Finland

## Abstract

An Atlantic salmon (*Salmo salar* L.) expressed sequence tag (EST) database consisting of 58 146 ESTs was screened for microsatellite sequences. Subsequent development of 75 polymorphic EST-associated microsatellite markers in this species is described together with cross-species amplification results of 133 gene-associated tandem repeat markers in five salmonid species (*Salmo trutta*, *Oncorhynchus mykiss*, *Salvelinus alpinus*, *Thymallus thymallus*, *Coregonus lavaretus*). The number of alleles among EST-linked microsatellites in Atlantic salmon ranged from two to 41 with an average of 12 alleles per locus. Cross-species amplification resulted in detection of a total of 111 polymorphic locus-species combinations (12–32 loci per species).

**Keywords:** cDNA, EST, expressed sequence tag, microsatellite, salmonidae, *Salmo salar*

Received 30 September 2004; revision received 11 November 2004; accepted 3 December 2004

The Atlantic salmon (*Salmo salar* L.) is a highly prized fish because of its widespread use in fisheries, recreational angling and aquaculture. It also serves as a prominent species in a wide range of research areas. Large-scale cDNA sequencing efforts have substantially expanded the amount of available information from the protein coding parts of the genome in salmonids during the recent years (e.g. Rise *et al.* 2004). Currently, over 91 000 Atlantic salmon expressed sequence tags (ESTs – short single-pass sequence reads from cDNA; representing its corresponding mRNA) are publicly available providing a rich source for screening for single nucleotide and insertion/deletion (indel) polymorphisms, as well as for developing gene-associated microsatellite markers from untranslated regions. Such highly variable markers therefore represent an attractive source of loci for the screening of selective footprints in the expressed part of the genome.

Here, we describe the EST database search and development of 75 EST-associated polymorphic microsatellite markers in the Atlantic salmon and report the transferability of 133 Atlantic salmon gene-associated primer-pairs in five species from the family Salmonidae (brown trout, *Salmo trutta*; rainbow trout, *Oncorhynchus mykiss*; Arctic charr, *Salvelinus alpinus*; European grayling, *Thymallus thymallus*; common whitefish, *Coregonus lavaretus*).

In total, 58 146 Atlantic salmon EST sequences from GenBank were screened for mono-, di-, tri- and tetranucleotide microsatellite repeats using TANDEM REPEATS FINDER (Benson 1999) with the following parameters: match 2; mismatch 7; indel 7; minimum alignment score 50. Microsatellites were detected in 1154 (2.0%) of the sequences. Because EST databases are highly redundant, microsatellite containing ESTs were clustered using CAP3 program with 40-bp overlap and 95% identity criterion in order to identify homologous loci (Huang & Madan 1999). Altogether, 144 contigs and 510 singletons were detected. Polymerase chain reaction (PCR) primers flanking 124 dinucleotide and seven tetranucleotide microsatellites with more than 11 and seven repeat units,

Correspondence: Anti Vasemägi, Fax: +4690 123 729; E-mail: anti.vasemagi@vabr.slu.se

respectively, were designed using PRIMER3 software (Rozen & Skaletsky 1998) with targeted amplicon size from 95 to 400 bases. Similarity searches with the microsatellite containing EST sequences for identifying gene homology were conducted using BLASTN and BLASTX with the default parameters (Altschul *et al.* 1990) and matches were considered to be significant when the probability was less than  $1 \times 10^{-4}$ . In seven cases, BLASTN search resulted in significant ( $P < 1 \times 10^{-4}$ ) matches against published Atlantic salmon genomic microsatellite sequences. However, these ESTs likely do not correspond to the already published microsatellite loci as significant similarities were restricted only for short stretches of the whole sequences. One mini- (*MHCII $\alpha$* ) and two microsatellite markers (*MHCI*, *TAP2B*) linked to genes associated with the immune response (Grimholt *et al.* 2002; Stet *et al.* 2002) were also included for cross-species amplification. An initial screen for amplification reliability was performed using four unrelated salmon specimens (each from a different European population) and 12 progeny from two families.

A 5'-GTTT 'tail' was added to every reverse primer to improve scoring quality (Brownstein *et al.* 1996). Each forward primer was 5' end-labelled with a fluorescent dye. PCR amplifications were performed using MJ Research PTC-100/200 thermal cyclers in 10- $\mu$ L volumes with *c.* 100 ng of DNA, 0.3  $\mu$ M of each primer, 0.2 units of BioTaq DNA polymerase (BioLine), 250  $\mu$ M of each dNTP, 1.5 mM of MgCl<sub>2</sub> and 1  $\times$  NH<sub>4</sub> reaction buffer [160 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 670 mM Tris-HCl pH 8.8, 0.1% Tween-20]. Initially, a touchdown PCR profile (TD:60–50) was used which consisted of an initial denaturation step of 2 min at 94 °C, followed by 20 cycles of 30 s at 94 °C, 30 s at 60 °C, 30 s at 72 °C, with the annealing temperature decreasing 0.5 °C per cycle, followed by 15 cycles of 30 s at 94 °C, 30 s at 50 °C, 30 s at 72 °C and a final elongation step was lengthened to 5 min. When the initial amplification was unsuccessful, an alternative touchdown PCR protocol (TD:55–45) was used where the initial annealing temperature 55 °C was lowered to 45 °C as described previously. All cross-amplification tests were performed using TD:55–45 protocol involving two individuals of each species. Electrophoresis was performed using MEGABACE 1000 96-capillary sequencer and allele sizing was conducted using MEGABACE GENETIC PROFILER version 2.0 software (Amersham Biosciences) followed by manual corrections.

Altogether, 75 EST loci that gave high quality amplification products were selected for further genotyping of 196 individuals from six anadromous and two landlocked salmon populations in northern Europe. Of these 75 ESTs, 31 showed significant ( $P < 0.001$ ) similarities to other sequences in GenBank (Table 1). The number of alleles per locus ranged from two to 41 with an average of 12 alleles per locus. No loci were found to be monomorphic indicating that the threshold value for minimum number of tandem

repeats (11 and seven uninterrupted di- or tetranucleotide repeats, respectively) could possibly be decreased when screening for polymorphic microsatellites in Atlantic salmon. Fifteen markers revealed a significant ( $P < 0.05$ ) heterozygote deficiency before correction for multiple testing (*c.* five significant  $P$  values are expected to occur by chance alone) using the Markov chain method implemented in GENEPOP 3.1 (Raymond & Rousset 1995). Information of all 133 tested primer pairs are available at the journal website (Appendix I). Electrophoregrams of the 78 markers that gave high quality amplification products generated using software MEGABACE GENETIC PROFILER version 2.0 are available at <http://www.helsinki.fi/~primmer/>. Exact tests for linkage disequilibrium between loci within eight populations gave 1314 significant  $P$  values ( $P < 0.05$ ; uncorrected for multiple testing) out of 35 093 tests using the Markov chain method implemented in GENEPOP 3.1 (Raymond & Rousset 1995). For majority of cases, linkage disequilibrium (LD) did not involve the same pair of loci. However, significant ( $P < 0.05$ ; uncorrected for multiple testing) LD was observed three times in seven locus pairs (CA053162 & CA057681; CA060177 & CA056586; CA040282 & CA046540; CA054978 & CB515794; CA058128 & TAP2B; MHCI & TAP2B; CA058586 & Ssa422) and four times in three locus pairs (CA767838 & CA058557; CA062068 & Ssa197; Ssa422 & Ssa289).

Cross-species amplification in five salmonid species revealed the highest number of polymorphic loci with little or moderate stutter in brown trout and rainbow trout (32 loci; Fig. 1). The number of polymorphic markers detected in Arctic charr, European grayling and common whitefish was lower (22, 12 and 13 loci, respectively with little or moderate stutter). These are likely very conservative polymorphism estimates because only two individuals of each species were assessed. Nevertheless, the number of polymorphic markers in the tested species is generally negatively correlated with their evolutionary divergence from Atlantic salmon. Detailed cross-amplification results of 133 gene-associated markers are reported at the journal website (Appendix II) and at <http://www.helsinki.fi/~primmer/>.

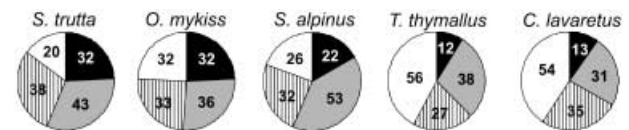


Fig. 1 Cross-species amplification summary results in five salmonid species shown as pie diagrams with the number of markers classified to each category. Black, polymorphic, little or moderate stutter; grey, monomorphic, little or moderate stutter; striped, considerable stutter and/or multiple bands and/or weak amplification; white, no amplification.

**Table 1** Characterization of 75 expressed sequence tag (EST) linked microsatellite markers in the Atlantic salmon and their amplification in five other salmonid species. Additional details of all 133 tested gene-associated markers can be found in the online Appendices I and II (<http://www.helsinki.fi/~primmer/>)

Accession no./ Locus name	Motif	Temp.	Size		$H_E$	$H_O$	HWE	Homology/Identity	Primer 5'–3'	Cross-species amplification				
			A	range (bp)						S.t.	O.m.	S.a.	T.t.	C.l.
CA046540	(AT) <sub>24</sub>	TD:60–50	18	314–348	0.86	0.69	0.024	<i>O. mykiss</i> vitellogenin receptor (vtg receptor gene)	F: TCCACTATATGATCACAATAACCTTTT R: <u>GTTTCCAGACACA</u> AAATGAGAGACC	?	?	+	?	P
CA048828	(CA) <sub>19</sub>	TD:60–50	24	251–307	0.73	0.59	ns	–	F: GAGGGCTTCCCATACAACAA R: <u>GTTTAAGCGGTGAGTTGACGAGAG</u>	P	+	+	0	P
CA048687	(AG) <sub>19</sub>	TD:60–50	2	203–205	0.05	0.04	ns	<i>D. rerio</i> Rh blood group-associated glycoprotein (rhag)	F: CAGAGACAGAGGGTCAGCCTA R: <u>GTTTCCCATCATCGTAGTCCACA</u>	+	P	+	0	?
CA048302	(AC) <sub>20</sub>	TD:60–50	23	201–255	0.93	0.74	< 0.001	–	F: TTGCCACCTCTAAACGCTTC R: <u>GTTTAAATGAACCCAGCCATACA</u>	P	0	0	+	0
CA048136	(CT) <sub>12</sub>	TD:60–50	4	347–355	0.57	0.39	ns	–	F: CCACATCAACAAGGCAAGAA R: <u>GTTTGATTGACAGAGAGCGGTGA</u>	0	0	?	0	P
CA047718	(AT) <sub>31</sub>	TD:55–45	13	242–288	0.74	0.61	ns	<i>O. mykiss</i> atrial natriuretic peptide (anp)	F: ATTTACCCGCTGGTGTATGTC R: <u>GTTTGCAAAGCCCTCATGTTGATT</u>	+	P	0	+	?
CA047634	(GT) <sub>13</sub>	TD:60–50	6	248–262	0.73	0.66	ns	–	F: GCTCTCACACAGACGCTGAC R: <u>GTTTCTGACTTGAGACGTCGGACA</u>	+	0	0	0	0
CA047220	(AT) <sub>13</sub>	TD:60–50	9	357–373	0.62	0.57	ns	–	F: AGCGTTTACGTCGAATCCAA R: <u>GTTTCTCATGGAGGGTGGAAAGTGT</u>	0	+	?	?	?
CA047146	(AT) <sub>29</sub>	TD:55–45	24	306–368	0.79	0.58	ns	–	F: AACAGAGATGGAACACAGCA R: <u>GTTTCCCAATCTAGCAGGGGAAAT</u>	+	+	P	0	0
CA060208	(CA) <sub>19</sub>	TD:60–50	11	186–214	0.76	0.5	ns	–	F: GCAACAATTCCTTTTGACC R: <u>GTTTCGTGCAGTAGGAAAGGGGTA</u>	P	0	0	0	0
CA044002	(GT) <sub>25</sub>	TD:60–50	9	205–223	0.77	0.57	ns	<i>O. mykiss</i> carbonyl reductase/20beta-hydroxysteroid dehydrogenase B gene	F: TTCTTGAGATGCCACACCTG R: <u>GTTTGTGTTCCTCATGTTTCACGA</u>	?	?	?	?	?
CA043931	(GT) <sub>14</sub>	TD:60–50	17	142–202	0.63	0.51	0.003	–	F: TGTGTGTGCATGTGTCCTTG R: <u>GTTTGTCCCTGCTGTTTCCACTCT</u>	?	?	+	?	?
CA042613	(GT) <sub>21</sub>	TD:60–50	14	107–149	0.71	0.64	ns	–	F: GCCAAGTGTCTTTCCTGTGAAA R: <u>GTTTCAGTCCACCTCGGAAAATC</u>	+	P	P	0	+
CA041953	(TA) <sub>15</sub>	TD:60–50	13	140–166	0.76	0.69	ns	–	F: TCACAGTCCGAGGAGTAAG R: <u>GTTTGTGACCCAGAGACAGATGACCT</u>	+	P	+	+	0
CA042465	(AT) <sub>12</sub>	TD:60–50	7	236–248	0.59	0.49	ns	–	F: CAGGAAAAGCCCATGAGAC R: <u>GTTTCATTGTTTCATAACGCACCA</u>	+	P	+	+	+
CA061261	(CA) <sub>22</sub>	TD:55–45	19	286–332	0.87	0.34	< 0.001	–	F: GGAGTATCAGGTGGGCAAA R: <u>GTTTGGGAGACCGTPTTGGTATTA</u>	+	+	0	0	?
CA039983	(GT) <sub>13</sub>	TD:60–50	6	269–279	0.52	0.48	ns	–	F: GCGGCCCTTAGTGTAATCAA R: <u>GTTTCTCGCCAGTCACTCTTCAA</u>	+	P	?	?	?
CA039543	(AT) <sub>26</sub>	TD:60–50	41	158–250	0.95	0.66	< 0.001	–	F: CACACAAGATTGGATTGAGCTT R: <u>GTTTCTCCCTTGTPTTTCCCAAT</u>	P	+	?	0	0
CA039240	(AT) <sub>26</sub>	TD:60–50	2	135–137	0.25	0.18	ns	<i>O. mykiss</i> GalNAc alpha 2.6-sialyltransferase (rtSt6GalNAc)	F: GCTGACCTAAATAACAGTPTTGGTGT R: <u>GTTTGGAAATCCCTCAAACCCCTTC</u>	?	?	+	0	P

Table 1 Continued

Accession no./ Locus name	Motif	Temp.	A	Size range (bp)	$H_E$	$H_O$	HWE	Homology/Identity	Primer 5'–3'	Cross-species amplification				
										<i>S.t.</i>	<i>O.m.</i>	<i>S.a.</i>	<i>T.t.</i>	<i>C.l.</i>
CA038592	(AT) <sub>12</sub>	TD:60–50	24	340–426	0.89	0.81	ns	–	F: AAGCATCAAACCAACCTCATT R: <u>GTTTCGGGGTGAAGATGCTACT</u>	+	0	P	P	0
CB515794	(GT) <sub>26</sub>	TD:60–50	16	265–309	0.84	0.71	ns	<i>G. gallus</i> similar to Claudin-12	F: CTCAGTGCATGTCTCCAAC R: <u>GTTTCATCCTGTCTCTGCTGACTG</u>	P	?	P	?	+
CB514761	(AC) <sub>26</sub>	TD:55–45	22	144–202	0.73	0.57	0.002	<i>M. musculus</i> Ariadne-1 protein homolog (ARI-1)	F: CTGTCCTTGGGCACATTTT R: <u>GTTTACAGCTCTGGTTCCGACA</u>	0	+	+	+	0
CA045807	(AC) <sub>12</sub>	TD:60–50	3	138–144	0.11	0.12	ns	<i>D. rerio</i> clone DKEY-29 N1 in linkage group 7	F: GGGACTCAGATAACGGCAAA R: <u>GTTTGTAAATATGGGCGTGTGTGA</u>	?	?	?	?	?
CB514369	(AG) <sub>12</sub>	TD:60–50	2	197–199	0.17	0.15	ns	–	F: CAATCATGCCTTCAGACTGTG R: <u>GTTTCCCATTTCTCCCAATCAC</u>	+	0	?	0	?
CB514068	(AC) <sub>14</sub>	TD:60–50	9	206–222	0.77	0.69	ns	–	F: TCTCAAATTCATACAATGTACAAGC R: <u>GTTTGGGAAAGGAACAGGAGAA</u>	?	0	0	0	0
CB512797	(AC) <sub>13</sub>	TD:60–50	20	268–436	0.58	0.47	0.048	<i>S. salar</i> for synapsin2 (SYN2)	F: GGACGAGGACCCTCCAAT R: <u>GTTTGGGGTGTGCTGAGGAGTATTT</u>	P	+	P	+	+
CB512617	(GT) <sub>12</sub>	TD:60–50	2	198–200	0.11	0.09	ns	–	F: GACCCGATTAGGCAGGACTT R: <u>GTTTGAGCAACTGTCCAGAGAGG</u>	P	?	?	?	?
CB517778	(GA) <sub>13</sub>	TD:60–50	6	398–408	0.65	0.61	ns	<i>T. nigroviridis</i> full-length cDNA	F: CCCAGCTGAGGCTCTTTATG R: <u>GTTTCTCCACATGTTCTGTC</u>	+	+	+	+	0
CB517710	(TA) <sub>17</sub>	TD:60–50	6	132–148	0.28	0.27	ns	–	F: GGATATGTGTGGCGCTCTGT R: <u>GTTTCCATTTCTCCATCCTTCCTT</u>	+	?	+	+	0
CB517044	(TA) <sub>21</sub>	TD:55–45	28	351–425	0.87	0.82	ns	–	F: CACCAAGCATGGGAAGCTAT R: <u>GTTTGTCTGCCACACAGGCTACTTT</u>	0	0	+	0	0
CA060177	(TGAG) <sub>18</sub>	TD:60–50	27	294–374	0.91	0.83	ns	–	F: CGCTTCCTGGCAAAAATTA R: <u>GTTTGAGCACACCCATTCCTCA</u>	P	0	+	0	+
CA767838	(AT) <sub>18</sub>	TD:60–50	5	214–224	0.63	0.5	0.007	<i>O. mykiss</i> clone pOT53m4006u19	F: GCAGGAGAGATCGAGAACA R: <u>GTTTGTGACTGCTGGCTGATG</u>	+	+	+	0	+
CA050122	(AG) <sub>12</sub>	TD:60–50	4	149–159	0.32	0.29	ns	<i>O. mykiss</i> interleukin-1 beta (IL1B1S)	F: TCCAAGCGTTCATGTGTTA R: <u>GTTTCCCAGTCTCCCTCTGCTTAG</u>	+	+	0	0	+
CA769358	(AC) <sub>22</sub>	TD:60–50	18	111–163	0.65	0.36	< 0.001	<i>O. mykiss</i> VHSV-induced protein-5	F: TGACGCCATATGCAAAGAGA R: <u>GTTTCTTTGTCTGCAAAACGTGA</u>	P	0	0	0	0
CA064581	(GT) <sub>16</sub>	TD:60–50	6	119–135	0.64	0.55	ns	<i>G. gallus</i> putative scavenger receptor MARCO	F: GATGCAACAAGTCTCAATACCA R: <u>GTTTCCATCGAAAAGACTGGAGGA</u>	?	0	?	+	0
CA064333	(AT) <sub>16</sub>	TD:60–50	19	138–186	0.84	0.63	ns	–	F: CATGGGTTTGTGTCCAGGATG R: <u>GTTTGTGATACGGGATTTTTCC</u>	?	0	+	?	?
CA064557	(GA) <sub>19</sub>	TD:60–50	6	345–361	0.30	0.29	ns	<i>S. salar</i> microsatellite Rsa180	F: CACAGGCACACTCCTCAT R: <u>GTTTCAGGTGAAGAGCATGACCAA</u>	?	0	+	+	P
CA061687	(GT) <sub>14</sub>	TD:60–50	3	251–255	0.45	0.39	ns	<i>S. salar</i> Ran protein	F: ACAGCGAGCGAACAAGTCT R: <u>GTTTCTTCAGTGTCTGGGGTGT</u>	+	?	+	?	?
CA061021	(GT) <sub>12</sub>	TD:60–50	3	98–102	0.36	0.31	ns	–	F: TCACAACAAGAAGTCCCAGATG R: <u>GTTTCTCCCTGACACTCTCGT</u>	?	P	+	P	+
CA039588	(AT) <sub>13</sub>	TD:60–50	3	90–96	0.30	0.25	ns	<i>O. mykiss</i> G-protein (P-ras)	F: TTGAGTAAACGTAAAGGACAGTCTTC R: <u>GTTTGTGTTGTCTGCTGTGTGATG</u>	+	0	?	+	?
CA058000	(AT) <sub>12</sub>	TD:60–50	10	276–294	0.73	0.62	ns	–	F: CAAACTCCCGCTCTCTCTA R: <u>GTTTGACCCCTTCCCAATTAGTGT</u>	+	P	+	0	0

Table 1 Continued

Accession no./ Locus name	Motif	Temp.	A	Size range (bp)	$H_E$	$H_O$	HWE	Homology/Identity	Primer 5'-3'	Cross-species amplification				
										<i>S.t.</i>	<i>O.m.</i>	<i>S.a.</i>	<i>T.t.</i>	<i>C.l.</i>
CA053293	(AC) <sub>13</sub>	TD:60-50	5	151-163	0.71	0.59	ns	—	F: TCTCATGGTGAGCAACAAACA R: <u>GTTT</u> ACTCTGGGGCATTTCATTTCAG	P	P	+	+	+
CA040282	(AT) <sub>24</sub>	TD:60-50	11	125-147	0.78	0.74	ns	—	F: TGCAAGTAAAGGCAGGGTTT R: <u>GTTT</u> GTGGTAGGATTGGGGTTCCT	P	P	?	+	+
CA040580	(AT) <sub>12</sub>	TD:60-50	5	114-124	0.59	0.45	ns	<i>O. mykiss</i> Wilms' tumour suppressor 1a (WT-t1a)	F: TCAATGGGGAACAATAACAACA R: <u>GTTT</u> GGTCTGTTCCCTCTGTTTT	?	+	+	+	P
CA047944	(AC) <sub>13</sub>	TD:55-45	7	276-288	0.47	0.26	ns	—	F: GCCGCCAGATTATCAGTAA R: <u>GTTT</u> GTTCCTCAACCAAAAACTGAA	+	+	0	0	+
CA053162	(AC) <sub>17</sub>	TD:55-45	12	332-372	0.63	0.41	< 0.001	—	F: CTCAACTCGTTCCTTCTCG R: <u>GTTT</u> CAGAAAATCCTTTTGGCTCA	+	0	+	0	0
CA054538	(AC) <sub>14</sub>	TD:60-50	6	108-118	0.50	0.39	ns	<i>D. rerio</i> Cyclin E	F: AGCTACTGGTCCCAAACT R: <u>GTTT</u> AAGGTGGACTTGGCTTGATG	+	P	+	+	0
CA054957	(GT) <sub>15</sub>	TD:60-50	8	219-239	0.66	0.57	ns	<i>O. mykiss</i> Cu/Zn-superoxide dismutase (SOD1)	F: GGGTCATTTGGGACACAGTT R: <u>GTTT</u> GGAGACAACGAGGAGATCG	?	P	?	0	?
CA055873	(TG) <sub>12</sub>	TD:60-50	3	347-351	0.42	0.36	ns	<i>S. salar</i> CHA(6.3) gene	F: GCACAGAGCTGGGTTTCAGTA R: <u>GTTT</u> GTGTTTGGGAGACTGGTCA	P	+	+	?	?
CA063143	(CA) <sub>21</sub>	TD:60-50	12	190-216	0.71	0.62	ns	—	F: CATGAACATCGAGCAGGCTA R: <u>GTTT</u> AGCTCATTTCTCCACGTTTGA	+	?	0	+	+
CA058128	(AC) <sub>12</sub>	TD:60-50	6	144-156	0.64	0.54	ns	<i>D. rerio</i> ATPase. H + transporting. V1 subunit G isoform 1	F: GCACTGTATTTGGTCTCCACA R: <u>GTTT</u> CTGTTTCCCATCTTCATTGC	+	+	+	+	+
CA058580	(GT) <sub>12</sub>	TD:60-50	4	199-209	0.46	0.46	ns	—	F: ATAACATGCAAGCGGTTTCC R: <u>GTTT</u> GTGTGGAAGTGTGAGTTGC	+	P	+	?	?
CA059521	(TA) <sub>15</sub>	TD:60-50	3	146-158	0.48	0.36	0.012	<i>D. rerio</i> COP9 constitutive photomorphogenic homolog subunit 7A	F: GCGCTCAAGTCTGTCACTCA R: <u>GTTT</u> TCGATAACGTCACCAATGCAC	+	+	+	P	+
CA056586	(AG) <sub>14</sub>	TD:60-50	11	366-388	0.77	0.6	ns	<i>O. gorbuscha</i> microsatellite locus Ogo2	F: GCCTACATCGCACACCATAA R: <u>GTTT</u> CCACTCATTTCTCGCTTTTCA	+	P	0	?	?
CA038562	(GT) <sub>20</sub>	TD:60-50	8	139-157	0.62	0.53	ns	<i>D. rerio</i> proto galectin (Gal1-L2)	F: TAGTGGCTCCATCCATTGGT R: <u>GTTT</u> CATTGGCTTTCCAGAGGTC	?	P	?	P	+
CA058557	(AT) <sub>17</sub>	TD:60-50	8	188-208	0.75	0.62	0.009	—	F: TCCAACACCACAATCTTCAAGT R: <u>GTTT</u> GTGCGTTTCGGGTGTAATG	+	+	+	P	P
CA037514	(CT) <sub>12</sub>	TD:60-50	3	280-286	0.18	0.16	ns	<i>S. salar</i> peroxisome proliferator activated receptor gamma (PPAR gene)	F: TGAACATCGGCAGTACCTGA R: <u>GTTT</u> TATGAGGGAGGAAAGAAGGA	?	?	+	0	0
CA062844	(CA) <sub>13</sub>	TD:60-50	5	189-201	0.70	0.6	ns	—	F: TGACACTGTGGCTGTCTCT R: <u>GTTT</u> GAGTTCTGGGTTATTTATTTCACA	+	+	+	0	+
CA057681	(CA) <sub>15</sub>	TD:55-45	6	357-375	0.15	0.16	ns	—	F: CTATTGTCTCCCGATGTGT R: <u>GTTT</u> CCATCCAGGCTTCTTATTTCACA	0	P	+	0	+
CA062621	(TA) <sub>17</sub>	TD:60-50	11	348-372	0.68	0.55	ns	—	F: TTAAACTCCTGCCCTGTGG R: <u>GTTT</u> CTTCCAAGGCTTGATGTCC	0	+	?	P	+
CA062068	(CT) <sub>13</sub>	TD:60-50	10	227-249	0.69	0.62	ns	—	F: TCATCAAGGCTTTGTGTCAG R: <u>GTTT</u> GTGCCAAATTATTTGGCCGTA	+	P	+	0	0
CA059136	(TA) <sub>22</sub>	TD:60-50	27	318-380	0.93	0.85	ns	—	F: AGGGTAGTGAGAAAGCAGCAA R: <u>GTTT</u> AACTGGCTGGCCATAGG	P	P	+	P	P

Table 1 Continued

Accession no./ Locus name	Motif	Temp.	Size range (bp)		$H_E$	$H_O$	HWE	Homology/Identity	Primer 5'–3'	Cross-species amplification				
										<i>S.t.</i>	<i>O.m.</i>	<i>S.a.</i>	<i>T.t.</i>	<i>C.l.</i>
CA058902	(TA) <sub>35</sub>	TD:60–50	2	180–182	0.35	0.26	0.011	–	F: CCAGCCAGAGAGGAACAGAC R: <u>GTTT</u> GATAGCCCATCACCAACC	P	P	+	P	?
CA058902	(TA) <sub>35</sub>	TD:60–50	34	201–271	0.84	0.67	0.028	–	F: CCAGCCAGAGAGGAACAGAC R: <u>GTTT</u> GATAGCCCATCACCAACC	P	P	+	P	?
CA058586	(AT) <sub>16</sub>	TD:60–50	22	202–278	0.65	0.26	< 0.001	<i>O. cuniculus</i> glycogen debranching enzyme	F: CCACACAAGTTTGACTGGA R: <u>GTTT</u> GGGTGAATGTCCCTTCTT	?	?	P	0	+
CA055420	(TGTA, TG) <sub>18</sub>	TD:60–50	21	356–442	0.63	0.51	ns	<i>Rattus</i> sp. transmembrane protein	F: AGGCAGTATTTGGCGACATC R: <u>GTTT</u> GTGCTGCCCTTATGTCCTG	0	+	+	0	+
CA054978	(AC) <sub>22</sub>	TD:60–50	12	314–330	0.57	0.5	ns	–	F: ACACAACCCAGACACCAACA R: <u>GTTT</u> TCTGCCCTTCTGTCTTAA	+	+	0	?	0
CA054565	(AG) <sub>19</sub>	TD:60–50	4	112–126	0.38	0.36	ns	–	F: TCTGTGGTTCCCGATCTTTC R: <u>GTTT</u> CAACATTTGCCCTAGCCCGAGA	P	?	P	0	0
CA053480	(AC) <sub>15</sub>	TD:60–50	16	254–290	0.84	0.71	ns	–	F: TGGTCACAACCAAAATGGAA R: <u>GTTT</u> CCACTCCAGGGTGTCTGTA	?	?	P	?	?
CA055301	(CA) <sub>29</sub>	TD:60–50	21	217–263	0.87	0.8	ns	–	F: AGAACCAAGGGTACCGATCC R: <u>GTTT</u> GGGAAATGGGTGGTAAGAAAA	+	+	P	P	+
CA052545	(CA) <sub>47</sub>	TD:60–50	3	257–263	0.03	0.02	ns	–	F: CCAAAGGCTCGCTAAAAACA R: <u>GTTT</u> CAGTGACAGCATGAGGACAAA	?	0	+	0	0
CA050657	(TA, CA) <sub>13</sub>	TD:55–45	12	102–126	0.85	0.74	ns	<i>S. fontinalis</i> steroidogenic acute regulatory protein (StAR)*	F: GGTTTCTGGGTCAAAGGAA R: <u>GTTT</u> GAAATGAGTGGCTGTCCAAA	P	0	+	0	0
CA050376	(GT) <sub>17</sub>	TD:60–50	10	301–319	0.69	0.64	ns	–	F: AGAGAACCATGAGGGGGAAC R: <u>GTTT</u> CTTTCTGATGTGGGATGGA	P	P	P	0	?
BG935488	(CAAT) <sub>23</sub>	TD:60–50	18	166–234	0.92	0.83	ns	<i>O. mykiss</i> neutrophil chemotactic factor (LECT2)	F: TGACCCACCAAGTTTTTCT R: <u>GTTT</u> AAACACAGTAAGCCCATCTATTG	P	+	+	0	+
BG934281	(TCTG) <sub>14</sub>	TD:60–50	27	193–267	0.88	0.8	ns	<i>T. nigroviridis</i> unnamed protein*	F: ACTGCTTCTCCCTGTACA R: <u>GTTT</u> GCGAACCACACATATACCAC	?	0	0	?	?

TD:X-Y refers to touchdown PCR protocol (see text); A, number of alleles;  $H_E$ , expected heterozygosity;  $H_O$ , observed heterozygosity; HWE, significance value from Hardy–Weinberg equilibrium test uncorrected for multiple testing; ns,  $P > 0.05$ .

Underlined sequences indicate added 5'-GTTT 'tail' (see text).

References: \*Vasemägi *et al.* in press; cross-amplification results of *Salmo trutta* (*S.t.*), *Onchorhynchus mykiss* (*O.m.*), *Salvelinus alpinus* (*S.a.*), *Thymallus thymallus* (*T.t.*), *Coregonus lavaretus* (*C.l.*): P, denotes polymorphism, little or moderate stutter; +, indicates single band with little or moderate stutter; ? indicates considerable stutter and/or multiple bands and/or weak amplification; 0, indicates absence of PCR product.

## Acknowledgements

We are indebted to Leena Laaksonen for her excellent laboratory work. This work was supported by grants from KSLA, the Oscar and Lili Lamm's Foundation, the Centre of International Mobility, the Academy of Finland, the Finnish Ministry for Agriculture and Forestry and a NorFA mobility scholarship.

## Supplementary material

The following material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/MEN/MEN902/MEN902sm.htm>

**Appendix S1** One hundred and thirty-three gene-associated markers screened in Atlantic salmon.

**Appendix S2** Cross-amplification details of one hundred and thirty-three gene-associated markers in five salmonid species.

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573–580.
- Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques*, **20**, 1004–1010.
- Grimholt U, Drabløs F, Jørgensen SM, Høyheim B, Stet RJM (2002) The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. *Immunogenetics*, **54**, 570–581.
- Huang XQ, Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Research*, **9**, 868–877.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): a population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Rise ML, von Schalburg KR, Brown GD *et al.* (2004) Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Research*, **14**, 478–490.
- Rozen S, Skaletsky HJ (1998) PRIMER3. Available at [http://www-genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/genome_software/other/primer3.html).
- Stet RJM, de Vries B, Mudde K *et al.* (2002) Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. *Immunogenetics*, **54**, 320–331.
- Vasemägi A, Gross R, Paaver T, Koljonen M-L, Säisä M, Nilsson J (in press) Analysis of gene-associated tandem repeat markers in Atlantic salmon (*Salmo Salar* L.) populations: implications for restoration and conservation in the Baltic Sea. *Conservation Genetics*.