Molecular Ecology Notes (2004) **4**, 89–92 doi: 10.1046/j.1471-8286.2003.00578.x

PRIMER NOTE

Isolation, characterization and cross-salmonid amplification of 31 microsatellite loci in the lake whitefish (*Coregonus clupeaformis*, Mitchill)

SEAN M. ROGERS, MARIE-HÉLÈNE MARCHAND and LOUIS BERNATCHEZ
Québec-Océan, Département de biologie, Université Laval, Sainte-Foy, Québec, Canada, G1K 7P4

Abstract

Coregonine fish represent the most successful evolutionary lineage of salmonids with *Coregonus* as the most speciose salmonid genus inhabiting numerous postglacial lakes across the northern hemisphere. We isolated and characterized 31 polymorphic microsatellite loci in *Coregonus clupeaformis* with an average number of 5.3 alleles per locus (range three to eight) and an overall expected heterozygosity of 0.74 ± 0.11. Two loci revealed significant linkage associations through analyses of mapping families. Six additional salmonid taxa assessed for cross-species amplification revealed between 18 and 26 positive amplifications and between two and 12 polymorphic loci per species.

Keywords: *Coregonus*, cross-species amplification, microsatellite, salmonid, whitefish

Received 24 September 2003; revision accepted 14 November 2003

Coregonine fish represent the most successful evolutionary lineage of salmonids and consist of three genera, *Prosopium*, *Stenodus* and *Coregonus*. Their broad circumpolar distribution in the northern hemisphere has a southern limit that parallels the maximum glacier advance during the Pleistocene. *Coregonus* is the most speciose salmonid genus with over 28 recognized species (Reshetnikov 1988). The lake whitefish (*Coregonus clupeaformis*) is unique as it consists of a species complex with reproductively isolated limnetic and benthic species pairs coinhabiting numerous lakes in eastern Canada. As such, these young species are noted to be unique systems within which evolution and speciation may be investigated (Rogers *et al*. 2001; Bernatchez 2004). Furthermore, lake whitefish have long been recognized as an important sport and commercial freshwater fishery resource (Hoyle & Mathers 2002). These factors have necessitated the development of genetic tools to assist in the study of their evolution and conservation. To this end, we have developed whitefish microsatellite markers for the purpose of integrated genetic mapping and population genetics approaches. Here, we present 31 new *C. clupeaformis* microsatellite loci and examine cross-species amplification of these loci in the pygmy whitefish (*Prosopium coulteri*), common whitefish (*C. lavaretus*), rainbow trout (*Oncorhynchus mykiss*), arctic charr (*Salvelinus alpinus*), cisco (*C. artedi*) and Atlantic salmon (*Salmo salar*).

Genomic DNA was extracted from a lake whitefish liver using standard proteinase K phenol–chloroform techniques (Sambrook *et al*. 1989) and pooled in a partial Sau3A1 digest. Fragments ranging from 400 to 900 bp were purified from a 1.7% agarose gel using DEAE paper (Sambrook *et al*. 1989). These fragments were then ligated into the *Bam*HI site of phosphatase-treated pUC18 ('Ready to go' kit; Pharmacia). Competent DH5 *Escherichia coli* cells (Gibco BRL) were transformed with the ligation products and grown on agar plates containing X-Gal and IPTG. Colonies were blotted on Hybond N+ nylon membranes (Amersham) which were hybridized with synthetic $(TC)_{10}$, $(TG)_{10}$ $(CAC)_5CA$, $CT(CCT)_5$, $CT(ATCT)_6$ and $(TGTA)_6TG$ probes labelled with the digoxigenin (DIG) oligonucleotide kit (Boehringer Mannheim). The DIG nucleic acid detection kit (Boehringer Mannheim) was used for detection. Among the 1600 clones of the partial library screened, 151 (~9.4%) were identified as positive. DNA from positives was extracted using the QIAprep plasmid DNA prep kit (Qiagen) and sequenced after cycle sequencing using Big Dye Terminator version 3.0 (Applied Biosystems, Inc.) using an ABI 377 automated sequencer. All of these clones contained microsatellites with 31 primer sets amplifying polymorphic and scorable products in *C. clupeaformis*. Characteristics of these primer pairs are described in Table 1.

Correspondence: Sean Rogers. Fax: (418) 656 2043; E-mail: sean.rogers@giroq.ulaval.ca

© 2004 Blackwell Publishing Ltd

Table 1 Characterization of 31 microsatellite loci in *Coregonus clupeaformis*

Locus ID/ GenBank	Repeat motif	Primer sequences (5′–3′)	T_a	*n*	Range	*a*	H_O	H_E	Map	Accession no.
*Cocl-Lav*1	CAT(GT_6)ACG CAT(GT_6)ATT	F: CGCAGGATTTTATCTGGACA R: AAGGAGACATCATCCGTTTCA	55	10	200–230	6	0.70	0.80		AY453196
*Cocl-Lav*4	CTA(CA_{13})CTG	F: TGGTGTAATGGCTTTTCCTG R: GGGAGCAACATTGGACTCTC	57	10	140–160	4	0.60	0.58	*	AY453197
*Cocl-Lav*5	AAA(CA_{14})GGC	F: ATGATGAGGCCCCAGGTAAT R: AAACAGCTATGCCATGATTCG	63	5	130–160	3	0.40	0.56		AY453198
*Cocl-Lav*6	TGA(GT_{22})ATC	F: GCCATCATCCTCCCAGGAAAC R: CAGGGAATCTGCACTGGAGC	60	10	130–150	8	1.00	0.80	*	AY453199
*Cocl-Lav*8	GAA(CA_{24})AGA	F: GCTGGAGCCACATGACATTA R: ATGTTTTTCCATTGCCCAGA	57	10	80–110	6	0.50	0.74	*	AY453200
*Cocl-Lav*10	TTT(GT_8)AGG	F: CAGTGGAGTTAATGAGTGCC R: GTGGAAATTGAATACTGCGG	57	10	250–270	4	0.60	0.51	*	AY453201
*Cocl-Lav*15	ACT(GT_3)AA(GT_{11})GAA	F: GCAGTCTCTGCTTATTGAGGAT R: GTGTCTGCATTCAGGTCACAGC	57	10	150–175	7	1.00	0.65		AY453202
*Cocl-Lav*18	AGC(GA_{12})GGT	F: AACAAACTAAAACATCCCAAGTC R: TTAGATTGGGGCCTACCTTG	57	10	145–160	6	0.70	0.73	*	AY453203
*Cocl-Lav*19	TTT(GT_{14})CAG	F: TCACTGTACAACAGAATAGGGAAA R: ATCCCTGATAAGCAGCCTCA	57	10	250–270	5	0.80	0.75	*	AY453204
*Cocl-Lav*22	CGT$(CA_{11})(CT_6)$CGC	F: GAGAGGGGGTATGTCTGT R: ATCGGAGTTTAGTAACCAC	51	9	110–125	5	0.78	0.73		AY453205
*Cocl-Lav*23	ATA(GT_8)CTT	F: GGGGAGGCAGTGGTGTATTA F: TGTGGAATTGTGAGCGGATA	57	9	175–200	8	0.78	0.76	*	AY453206
*Cocl-Lav*27	TTG(GT_6)GAT	F: TGACTCTTCCCCATTCATCC R: CCGAGAGGTGGAGAAAACAG	55	10	115–130	5	0.60	0.62		AY453207
*Cocl-Lav*28	ATA(CA_{24})CGC	F: ACAATAGCAGGCCATTCAGG R: CCAATCTTCAAAGCCATTTCA	60	9	170–190	6	0.56	0.80	*	AY453208
*Cocl-Lav*32	GGA$(GT_9)(CT_{23})$GCC	F: CCCCACGTCTCTCCCTTAAT R: CGCTGTCAACTTTCCCTCTC	60	9	250–320	5	0.44	0.75	*	AY453209
*Cocl-Lav*38	GAG(CA_{19})ACC	F: GCCATGATTACGATTTCGAG F: GCAGGACAGTAGTGTCTCCAT	55	10	110–128	5	0.70	0.62	*	AY453210
*Cocl-Lav*41	TTT(CT_{36})TTT	F: AAACAAACAGTGGTGGAGTGG R: GCCAGCACTCTCTCATGCTTTT	60	9	130–150	7	0.56	0.77		AY453211
*Cocl-Lav*45	TTT(GT_{13})CTC	F: GAGTGACAGCAGGGAGCAG R: GGCTCGGTTGAAAGTTGAGA	60	10	225–250	4	0.60	0.62		AY453225
*Cocl-Lav*49	ACT(GT_{17})GGG	F: AGCCAGTTGGAGGCTATTTG R: AGGGCTGCTGTTGAAGTCAT	55	10	164–180	5	0.80	0.74*		AY453212
*Cocl-Lav*49b		F: AGCCAGTTGGAGGCTATTTG R: AGGGCTGCTGTTGAAGTCAT	55	9	340–355	3	0.67	0.48		
*Cocl-Lav*52	GAT(GT_{52})TCT	F: GGCGATTGGGAGAGTGATTA R: ACAGAGCCCCAGATGGTAAC	55	9	160–175	5	0.67	0.70*		AY453213
*Cocl-Lav*61	AAA(CA_{24})AAA	F: CTCATGAGTAACATGATGCTTC R: GATCTTTACTGTCTGATTTTGTG	55	9	250–300	4	0.44	0.73*		AY453214
*Cocl-Lav*68	GGT(CA_{11})AAC	F: GTGTGTTACAAGTGGCTATG R: GTGATGGCTTTCAGAGGC	57	9	165–180	5	0.67	0.75*		AY453215
*Cocl-Lav*69	CTC(GT_9)TAG	F: CTCAACGTCGTCTGAGTG R: GTGTAATGACACTTCTCTGG	61	10	100–130	3	0.50	0.60		AY453216
*Cocl-Lav*72	ATA(GT_{23})GTC	F: CTCTCAAGATATCTAAGGAGG R: CGGAGTTTAGTAACCACATTG	60	10	162–194	6	0.60	0.80*		AY453217
*Cocl-Lav*74	TTT(GT_{26})CCC	F: GATCATAGTCTACAGATGG R: CCTCTGCTTCTCAACCAG	60	10	210–245	8	0.90	0.83		AY453218
*Cocl-Lav*80	AGT(GA_{14})AAG	F: GCATTTGTAAACCATCACC R: GAGGATCCACAAATAATTCG	55	7	170–196	7	0.71	0.84*		AY453219
*Cocl-Lav*216	ATT(GT_{29})TCT	F: CAGCGTTTGAATTGAGTTTC R: GTAGACAAAACCAATCAGG	55	5	240–270	5	0.60	0.54		AY453220
*Cocl-Lav*219	CAC(GT_{31})GAA	F: GAGATTACATTTCCTCATCC R: CCTCTAGTAGCTTGTGAC	55	8	150–160	4	0.75	0.61		AY453221

Table 1 *Continued*

Locus ID/ GenBank	Repeat motif	Primer sequences (5′–3′)	T_a	n	Range	a	H_O	H_E	Map	Accession no.
*Cocl-Lav*220	CGT(CA$_{15}$)TTT	F: GAAGCAGCTCTTATACACAC R: GAGCACACATGGTCCTTTAC	60	10	110–130	7	0.50	0.79		AY453222
*Cocl-Lav*221	TAA(CA$_{20}$)CGC	F: CAGGCAGCCATGAAGGTG R: GATCAGTTTACAGATGAGC	56	9	160–172	6	0.44	0.81		AY453223
*Cocl-Lav*224	AAA(CA$_{20}$)CGC	F: GTGGCAGGCAGCCATGAAG R: GACGTTAGTCACTGCTTTCC	60	10	230–260	4	0.60	0.58		AY453224

T_a, annealing temperature used in polymerase chain reaction; n, sample size of whitefish used to assess variability; range, allelic range (bp); a, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity. *Loci that have been segregated into lake whitefish linkage maps (see Results for linkage associations among loci). †Result of duplicated locus amplification, therefore sequence of repeat motif unavailable.

Microsatellite amplification was performed by extracting total genomic DNA from fin tissue using a standard phenol–chloroform procedure. The polymerase chain reaction (PCR) was performed in an 10-μL volume with 25–50 ng of DNA, 0.5 pmol each primer, 75 mM each nucleotide, 15 mM $MgCl_2$, 1 μl reaction buffer (100 mM Tris-HCl, pH 9, 500 mM KCl), 1 U of *Taq* polymerase (Applied Biosystems, Inc.) and 0.375 mM of dUTP TAMRA (Molecular Probe) fluorescent incorporation labelling. The PCR reactions were carried out in an Applied Biosystems 9700 DNA thermal cycler with an initial denaturation time of 3 min at 95 °C followed by 30 cycles of 95 °C for 30 s, 30 s at the locus-specific annealing temperature (see Table 1) and 60 s at 72 °C and 5 min of final elongation at 72 °C. The PCR products were separated by electrophoresis on 8% denaturing polyacrylamide gels. Bands were visualized on an FMBIO II scanner (Hitachi) with the number and size range of alleles scored using the GENESCAN-500 size standard (Applied Biosystems, Inc.). Locus variability was analysed using 10 samples of *C. clupeaformis* originating from East Lake (n = 3), Cliff Lake (n = 4) and Temiscouata Lake (n = 3) (Table 1; see Lu *et al*. 2001 for location details). The observed and expected heterozygosity were calculated using the software GENETIX 4.03 (Belkhir 2000).

The number of alleles among polymorphic loci ranged from three (*Cocl-Lav*5, *Cocl-Lav*49 and *Cocl-Lav*69) to eight (*Cocl-Lav*6, *Cocl-Lav*23 and *Cocl-Lav*84) with an overall average of 5.3 alleles per locus. Expected heterozygosity ranged from 0.51 (*Cocl-Lav*10) to 0.84 (*Cocl-Lav*80) with an overall expected heterozygosity among the 31 loci of 0.74 ± 0.11 (Table 1). As similar allele frequencies were observed among populations, we grouped all individuals (n = 10) in order to test for Hardy–Weinberg equilibrium using the Markov chain method implemented in GENEPOP version 3.1 (Raymond & Rousset 1995) to estimate the probability of heterozygote deficiency for each locus (Guo & Thompson 1992). One locus, *Cocl-Lav*221, revealed a significant heterozygote deficiency (P = 0.004). Null alleles may have contributed to the heterozygote deficiency observed for *Cocl-Lav*221 but more individuals will be needed to test this hypothesis. One microsatellite, *Cocl-Lav*49, was duplicate in the genome with the primer sequence amplifying two loci around 164 and 350 bp, respectively. Approximately half of these reported loci have been segregated into lake whitefish linkage maps dependent on locus polymorphism in parents of controlled crosses (Rogers, unpublished). Using LINKMFEX (Danzmann & Gharbi 2001), segregation results from pairwise comparisons of loci from two outbred whitefish families (n = 51 and 56) assessed linkage for these loci. Two loci (*Cocl-Lav*19 and *Cocl-Lav*23) were linked with significant likelihood odds ratio scores of 2.72 and 10.40 in the two families, respectively. Cross-species amplification using three individuals from each species (originating from different populations) revealed successful amplification (denoted by +) on average 68% of the time over all loci used. Polymorphic loci (denoted by ++) were found on average 20% of the time over all loci with *P. coulteri* and *C. lavaretus* amplifying the most polymorphic loci (39 and 35% of all loci, respectively) (Table 2). These results confirm the integrated amenability of these *C. clupeaformis* primers towards population and linkage genetic analyses for lake whitefish and closely related species of ecological and economical interest.

Acknowledgements

We would like to thank Gilbert Castilloux, Guillaume Côté, Christian Landry, Catherine Lippé, Makoto Matsuoka, Lucie Papillon, Madeleine Parent, Catherine Potvin, Robert St-Laurent and Manon Valcourt for technical assistance in the cloning and testing of primer sequences. We would also like to thank the subject editor for comments that improved the quality of the manuscript. This research was funded by NSERC and is a contribution to the Chair of conservation genetics.

Table 2 Cross-amplification data of six additional taxa using three individuals of each

Locus	*P. coulteri*	*C. lavaretus*	*C. artedi*	*O. mykiss*	*S. salar*	*S. alpinus*
*Cocl-Lav*1	++	+	–	+	+	+
*Cocl-Lav*4	++	++	+	++	++	–
*Cocl-Lav*5	–	–	–	+	–	–
*Cocl-Lav*6	++	+	++	–	–	+
*Cocl-Lav*8	+	++	++	–	–	–
*Cocl-Lav*10	++	+	+	++	+	++
*Cocl-Lav*15	–	+	+	+	+	–
*Cocl-Lav*18	++	++	–	+	+	–
*Cocl-Lav*19	–	–	–	–	–	–
*Cocl-Lav*22	–	–	–	–	+	–
*Cocl-Lav*23	–	+	–	+	++	+
*Cocl-Lav*27	++	++	+	+	+	+
*Cocl-Lav*28	–	++	–	++	+	–
*Cocl-Lav*32	++	++	++	–	+	+
*Cocl-Lav*38	++	++	++	+	–	+
*Cocl-Lav*41	+	+	–	+	–	+
*Cocl-Lav*45	+	+	+	+	–	+
*Cocl-Lav*49	–	–	+	+	–	+
*Cocl-Lav*49b	+	+	+	+	–	+
*Cocl-Lav*52	–	+	+	+	+	+
*Cocl-Lav*61	–	++	+	–	–	–
*Cocl-Lav*68	+	+	+	+	+	+
*Cocl-Lav*69	++	+	+	+	–	–
*Cocl-Lav*72	++	+	++	+	+	+
*Cocl-Lav*74	+	+	+	+	+	+
*Cocl-Lav*80	–	–	–	–	–	–
*Cocl-Lav*216	+	++	–	+	++	+
*Cocl-Lav*219	+	+	++	–	+	+
*Cocl-Lav*220	++	++	+	+	+	–
*Cocl-Lav*221	++	++	–	+	–	+
*Cocl-Lav*224	+	+	–	++	+	++
Total Amp	21	26	19	23	18	19
Total Poly	12	11	6	4	3	2

–, no amplification; +, amplification but insufficient data to determine polymorphism; ++, amplification and polymorphism; Total Amp, total number of + and ++ within species; Total Poly, total number of ++ within species.

References

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2000) GENETIX, *A WindowsTM Based Software for Population Genetic Analyses*. Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Bernatchez L (2004) Ecological theory of adaptive radiation: an empirical assessment from Coregonine fishes (Salmoniformes). In: *Evolution Illuminated: Salmon and Their Relatives* (eds Hendry AP, Stearns S), pp. 176–207. Oxford University Press, Oxford, UK.

Danzmann RG, Gharbi K (2001) Gene mapping in fishes: a means to an end. *Genetica*, **111**, 3–23.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Hoyle JA, Mathers A (2002) Lake Ontario commercial fishery. In: *Lake Ontario Fish Communities and Fisheries: 2001 Annual Report of the Lake Ontario Management Unit*, pp. 1–4. Queen's Printer for Ontario, Picton, Ontario, Canada.

Lu G, Basley D, Bernatchez L (2001) Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, **10**, 965–985.

Raymond M, Rousset R (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Reshetnikov YS (1988) Coregonid fishes in recent conditions. *Finnish Fisheries Research*, **9**, 11–16.

Rogers SM, Campbell D, Baird SJE, Danzmann RG, Bernatchez L (2001) Combining the analyses of introgressive hybridisation and linkage mapping to investigate the genetic architecture of population divergence in the lake whitefish (*Coregonus clupeaformis*, Mitchill). *Genetica*, **111**, 25–41.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.