MOLECULAR ECOLOGY
RESOURCES

Molecular Ecology Resources (2012) doi: 10.1111/j.1755-0998.2012.03163.x

Validation and comparison of microsatellite markers derived from Senegalese sole (*Solea senegalensis*, Kaup) genomic and expressed sequence tags libraries

M. J. MOLINA-LUZÓN, J. R. LÓPEZ, R. NAVAJAS-PÉREZ, F. ROBLES, C. RUIZ-REJÓN and R. DE LA HERRÁN

Departamento de Genética, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain

Abstract

In this work, we tested 100 potential new microsatellites (SSRs) equally derived from expressed sequence tag (EST) and enriched genomic-DNA libraries from Senegalese sole (*Solea senegalensis*, Kaup), a valuable cultured flatfish species. A final set of 69 new polymorphic microsatellites were validated after a population analysis, 37 of which corresponded to the first EST library constructed for Senegalese sole (EST-SSR). Although differences were not significant, EST sequences provided a higher proportion of quality markers (74%) than anonymous ones (64%). Most of the rejected anonymous SSRs (17 loci) were discarded because they did not generate PCR products; only one was monomorphic. On the contrary, all EST-SSRs gave PCR products, although monomorphism was more frequent (26%). Altogether, the number of alleles per locus was fairly similar in both SSR types, ranging from 2 to 19. The observed and expected heterozygosities varied from 0.105 to 1 and from 0.108 to 0.937, respectively. The main difference between the two sets was the percentage of annotated loci, being higher in EST-SSRs, as expected. Within the EST-SSRs, 46% of them showed flanking regions that significantly matched with EST sequences from other three flatfish species; however, the microsatellite itself was present only on half of these cases. These two new SSR sets constitute a suitable tool for fingerprinting, gene flow, genetic diversity, genome mapping studies and molecular-assisted breeding in this species.

Keywords: expressed sequence tag library, flatfishes, microsatellite, *Solea senegalensis*

Received 24 February 2012; revision received 27 April 2012; accepted 4 May 2012

Introduction

Senegalese sole is an economically valuable flatfish species of high interest for marine aquaculture mainly in Southern Europe. In fact, its culture has surged in the last few years in Portugal and Spain, contributing to the diversification of species with aquaculture potential in this region (Imsland *et al.* 2003). However, to date the culture of this species is not yet commercially successful, because of problems related mainly to feeding, mortality during the metamorphosis and the attack of bacterial diseases that cause severe losses and hamper production (Moriñigo *et al.* 2001; Imsland *et al.* 2003). Additionally, loss of genetic variability, which affects the capacity of fish to adapt to new conditions, is frequent in hatchery stocks (Sekino *et al.* 2003), causing inbreeding and thereby affecting progeny viability. In this context, the development of genetic maps and selective breeding programmes are commonly used to improve the culture of other aquaculture fishes. All these aspects make it necessary to characterize molecular markers in this species. Microsatellites, or simple sequence repeats (SSRs), have provided a powerful molecular tool to assess genetic variations, to assign parentage and to identify genomic regions related to productive traits, allowing selection based on molecular markers (MAS). In fact, SSR markers have been characterized in different cultured flatfishes such as Japanese flounder (Coimbra *et al.* 2003; Liu *et al.* 2006; Kang *et al.* 2008), turbot (Bouza *et al.* 2007), halibut (Reid *et al.* 2007), half-smooth tongue sole (Sha *et al.* 2011) and European plaice (Casas *et al.* 2009). However, while in such species a large number of microsatellite markers have been developed, allowing a thorough analysis of its genomes, only a few loci (about 40) have been reported in Senegalese sole to date (Funes *et al.* 2004; Porta & Alvarez 2004; Castro *et al.* 2006; Chen *et al.* 2008; De la

Correspondence: Roberto de la Herrán, Fax: +34 958244073; E-mail: rherran@ugr.es

© 2012 Blackwell Publishing Ltd

Herrán *et al.* 2008). It is thus necessary to develop a higher number of polymorphic markers for the construction of linkage maps, already developed in other flatfish species. Moreover, while expressed sequence tag (EST) databases have been used with high performance in other flatfishes, no microsatellite loci from gene sequences have been described in Senegalese sole. These EST-derived microsatellites have proved useful in comparative analyses because being present in coding regions, they are expected to be more conserved than are anonymous microsatellite loci. Here, we describe the first Senegalese sole microsatellite set obtained from EST sequences, together with a new set of anonymous SSR. Their characteristics, such as number of alleles and polymorphic degree and its conservation in other flatfish genomes, are described here.

Materials and methods

SSRs were detected using the software Perfect Microsatellite Repeat Finder (sgdp.iop.kcl.ac.uk/nikammar/repeatfinder.html). The criteria used to select both EST and anonymous SSR were as follows: at least eight repeats for dinucleotide and trinucleotide repeats and at least four repeats for tetranucleotide repeats. EST-SSRs were isolated from an established sequence tag library (Pleurogene project database; http://www.pleurogene.ca), whereas anonymous SSRs were obtained from previously described repeat-enriched genomic libraries (De la Herrán *et al.* 2008). Primers flanking the selected microsatellites were designed using the PRIMER 3 software (http://frodo.wi.mit.edu/primer3).

Microsatellite-marker polymorphism was determined using 28 wild individuals of Senegalese sole, used as breeders, collected from a fish farm located in the IFAPA Centre 'Agua del Pino' of Junta de Andalucía in Huelva, Spain. Previously, PCR conditions had been optimized by using template DNA from eight adult fish, varying the annealing temperature and the primer and $MgCl_2$ concentrations.

DNA was isolated using the phenol-chloroform procedure as described by Sambrook *et al.* (1989), and PCRs were finally performed in a 15-μL reaction mixture containing 1× buffer, 2 mM $MgCl_2$, 0.1 mM of each dNTPs, 1 U of *Taq* Polymerase (Biotools), 5 pmol of each primer set and approximately 25 ng of genomic DNA. PCR cycles were as follows: 5 min denaturation at 94 °C, 35 cycles of 1 min at 94 °C, 45 s at a primer-specific annealing temperature and 45 s at 72 °C. Finally, the products were extended for 7 min at 72 °C. The specific annealing temperature of each primer set is given in Tables 1 and 2. Polymorphism at each locus was screened using an ABI 3100 Avant sequencer (Applied Biosystems), by electrophoresis of PCR products on 6–8% denaturing polyacrylamide gels. Alleles were designated according to the PCR product size, which was determined using Gene Scan^{TM} 500 LIZ^{TM} Size Standard (Applied Biosystems) as a reference marker and the GENEMAPPER software (Applied Biosystems).

For each marker, the observed and expected heterozygosities, the polymorphism information content (PIC) value and the null-allele frequency (NAF) were all determined using the program CERVUS 3.0 (Kalinowski *et al.* 2006). Linkage disequilibrium analyses among the polymorphic maker pairs and deviations from Hardy–Weinberg (HW) equilibrium for each locus were estimated performing an exact test (probability test) in GENEPOP 4.0 (http://genepop.curtin.edu.au/) (Raymond & Rousset 1995a,b) using the Markov parameters provided by default. For annotation, microsatellite-containing sequences were BLASTed against the GenBank nucleotide database (http://www.ncbi.nlm.nih.gov/genbank/) by using BLASTX (*E*-value cut-off: $<10^{-3}$).

For the determination of the presence of Senegalese sole SSRs in other flatfish species, all EST sequences (and those anonymous sequences with a significant match on GenBank) were BLASTed using BLASTn (Altschul *et al.* 1990) against EST databases from Atlantic halibut (21 018 sequences), Japanese flounder (16 275 sequences) and turbot (6170 sequences), available from NCBI-EST database (http://www.ncbi.nlm.nih.gov/dbEST) except for turbot sequences, which were extracted from an specific-EST database (Pardo *et al.* 2008; Vera *et al.* 2011).

Gene ontology (GO) terms were assigned using BLASTX (*E*-value cut-off: $<10^{-3}$) as implemented by AmiGO v.1.8 software (http://www.geneontology.org/). In the annotated sequences, the presence of the repetitions out/in the translated region was determined for each SSR.

Results and discussion

One hundred potential SSR markers, 50 from genomic libraries and 50 from EST libraries, were characterized from sequences with accession numbers HE601639-HE601673 and HE600073-HE600122. Within those derived from genomic libraries, 34% failed to give a PCR product, 64% were found to be polymorphic among the tested fish and only 2% proved monomorphic. On the contrary, none of the 50 markers derived from EST libraries failed to give a PCR product; however, a higher degree of monomorphism (26%) was observed among them. The higher proportion of anonymous markers that failed to give a PCR product could be due to a presumably higher conservation of the EST sequences than genomic ones. Overall, a set of 69 new polymorphic microsatellites were validated after population analysis.

Tables 1 and 2 summarize anonymous- and EST-microsatellite features, respectively. Both sets are

© 2012 Blackwell Publishing Ltd

Table 1 Anonymous SSRs isolated from Senegalese sole with repeat motif, primer sequences, PCR annealing temperature (T), number of alleles (NA) and allele range in pair bases (pb), observed heterozygosity (H_O), expected heterozygosity (H_E), exact-probability test (*P*-value) and null-allele frequency (NAF)

Locus	Repeat motif in cloned sequenced	Primer sequences (5′–3′)	T (°C)	NA	Allele range (pb)	H_O	H_E	Polymorphism information content	*P*-value	NAF	Accession number
Polymorphic											
Mss1	$(TG)_{16}$	F: TGTCATTGAAGGGTGCACTAA R: AAACAACTTTTGCACGGTGA	58	11	188–216	0.7143	0.8481	0.815	0.0105	+0.0786	HE601639
Mss3	$(CA)_9$	F: ATTCTGTCCCCAATTCACCA R: AATTGTGGTCCGGGTTGTT	64.4	7	171–192	0.3462	0.7157	0.668	0.0033	**+0.3483**	HE601641
Mss5	$(CA)_{16}(CA)_6$*	F: AACTCCCAGCAGCTCCTACA R: CTCCCCATGAGGTTCAATGT	59	8	186–218	0.6667	0.6599	0.602	0.5825	−0.0003	HE601643
Mss7	$(CTAT)_4(CTAT)_9$*	F: GAGATGAGGGCTGAGACAGG R: ACCTGTGGAGACAGGCGTAG	64	19	122–212	0.8333	0.9193	0.893	0.0447	+0.0417	HE601645
Mss11	$(CT)_6(CA)_{10}$*	F: CATTAGGACGGGTCCATGTT R: TCATGTGGACTGGACCAGAA	58	9	206–222	0.8276	0.8826	0.853	0.4124	+0.0232	HE601646
Mss14	$(GA)_{17}$*	F: ACGTGAGAGGAAGTGGTGCT R: CGCCTCCAATGTCAGATTTT	60	11	183–207	0.6774	0.8646	0.834	0.0227	**+0.1181**	HE601648
Mss20	$(CA)_{15}$*	F: GCCGCATTCTAAACAGGTG R: CGGGTCTGTCAATCAAAGGT	63.5	4	174–204	0.4333	0.5774	0.488	0.6137	**+0.1143**	HE601651
Mss22	$(CCT)_{12}(TGC)_8$*	F: CGCCAGGTTGTTCAAACACT R: TTTGTCAGTCGTCCTCCAGA	60	4	165–174	0.5217	0.5729	0.463	0.8183	+0.0358	HE601652
Mss24	$(CA)_{16}$*	F: GGCTTCTGCTGCGTCTACTT R: CCCTGCTTTAGGGTGACAGA	63	8	156–182	0.7368	0.7425	0.689	0.6600	−0.0051	HE601654
Mss25	$(TTA)7(TTG)_5$*	F: GAGTGACTTCAACTTCGACCAA R: ACGGACACCAGGTTTGACTC	63	7	156–181	0.6129	0.7435	0.697	0.1714	+0.0973	HE601655
Mss27	$(CT)_{36}$*	F: CTCATCCTCCATTGCTCCTC R: ACTACCGTGGCGAGGTCAT	63	8	160–196	0.4583	0.4282	0.399	0.3433	−0.0498	HE601657
Mss28	$(GA)_8(CA)_{13}$*	F: TGCCCTGAACGATGACTGTA R: GAAATTTCCTCAGTAACCAAGAGG	63	9	239–267	0.6897	0.6400	0.610	0.8697	−0.0777	HE601658
Mss29	$(GA)_{18}$	F: TGGGAATAATGACAATGCAAA R: TTCCCTCACAGCATCATGTC	60	9	201–219	0.6000	0.8654	0.825	0.0010	**+0.1701**	HE601659
Mss30	$(ATG)_8$	F: GAACATGACGGAATCATGACA R: TCCCTGCCTTAATGACAGATAA	63	6	169–187	0.6500	0.6667	0.586	0.6780	+0.0002	HE601660
Mss32	$(AGG)_8$*	F: GGCACCAGCACTTTGATGTA R: GCCTGGGAATTATGACAACG	62	4	122–168	0.5217	0.5382	0.428	1′000	−0.0014	HE601661
Mss35	$(GTT)_{19}$	F: TTGATTTCCCCTCTTCCTCA R: TCAGACTGTGAAAGGTTGAAGG	57	11	129–191	0.7917	0.8014	0.764	0.7054	−0.0024	HE601662
Mss37	$(GT)_{10}$*	F: AAAGGCTGAATTAGCTTTGAACA R: GCATGACTCTGCCGTGACT	61	5	172–192	0.4167	0.5895	0.487	0.1592	**+0.1727**	HE601663
Mss42	$(CA)_{30}$	F: CCGAGTCCAGTTTATCACTGC R: TAGGCTGTCCACATGAATGG	63	15	102–170	0.6800	0.6841	0.660	0.5716	−0.0011	HE601665

© 2012 Blackwell Publishing Ltd

Table 1 (Continued)

Locus	Repeat motif in cloned sequenced	Primer sequences (5′–3′)	T (°C)	NA	Allele range (pb)	H_O	H_E	Polymorphism information content	*P*-value	NAF	Accession number
Mss44	$(CT)_{15}$*	F: TTGGCATGATTTGGCAGTT R: CAGTTGGGCAACCTATTATTTGA	62	10	226–248	0.7931	0.7834	0.752	0.4989	−0.0304	HE601666
Mss45	$(TG)_{11}$*	F: TGCAATGGTAGTTATTTAACATTTA R: TCCTTGCTCCCTTATTGCAT	48	17	150–408	0.8571	0.8393	0.811		−0.0221	HE601667
Mss46	$(CA)_{40}$*	F: TGCAGCTATGCAGATGTTGTT R: CTAGAGCCGCAGTTTGCAG	57	17	224–298	0.8000	0.9298	0.905	0.2802	+0.0656	HE601668
Mss47	$(CA)_{26}$*	F: TTCCTTGTCATTACAGAGGCAT CAATGTTAATGCACTGAGAAAGTT	60	17	120–295	0.7692	0.9027	0.875	0.2933	+0.0741	HE601669
Mss50	$(GA)_6(CA)_{20}$*	F: GCTGGGTTCGAGTTACAAGC R: GACTTGCTTATCCTTACATTCA	59.5	6	133–158	0.5833	0.6011	0.539	0.1850	−0.0050	HE601670
Mss53	$(GTT)_7$	F: CCCTGGGCTACAGACAATC R: GCTCCAGGGCTGCTATTGT	57	2	145–149	0.1111	0.1079	0.099	1′000	−0.0194	HE601640
Mss54	$(AG)_{16}$	F: GACCCTCGTAAGTGGGAGAA R: AGATTCCGCTGCAGTCCTT	57	2	166–168	0.4783	0.5072	0.373	1′000	+0.0184	HE601642
Mss55	$(CTT)_{11}$	F: TTTCAGATGACGAAGCAAACA R: CAAGGTGAGCAAACATGGA	57	4	109–119	0.3158	0.4054	0.368	0.2037	**+0.1320**	HE601644
Mss58	$(GT)_{11}$*	F: TGTAGTGCTCCCATTTCCTG R: GCGTCCCTTACTTCCTACCG	59.5	10	153–173	0.8636	0.8605	0.822	0.5996	−0.0174	HE601647
Mss63	$(CTAT)_{20}$*	F: ACATTTCATTTGCTGCTGCCACA R: GGGACATGTTGGCTGATTCT	61	16	112–206	0.9545	0.9366	0.909	0.8979	−0.0202	HE601664
Mss64	$(GAA)_{17}$*	F: CCTCGTCACTTTCACAGG R: GAAACACCAGTGAGCAGCAA	61	5	87–105	0.5000	0.4794	0.416	1′000	−0.0590	HE601650
Mss65	$(CAA)_{10}$*	F: TGGGATCAAATGAAGTCAGAAA R: TTGTGCAATATCACGAATGGA	57	2	137–140	0.0526	0.0519	0.050	1′000	−0.0055	HE601671
Mss66	$(AGAT)_{12}$*	F: ACTCTTTAACAAGTAAACCTGCATTA R: CATTTAACATGGATGAAACAGCA	54.2	13	106–178	0.8696	0.8097	0.768	0.6901	−0.0615	HE601672
Mss67	$(GA)_{12}$	F: CGGGTTCTGGTCTCATGTTT R: TCGTCTTCTGCTTCGACAAA	63	11	107–155	0.8846	0.8296	0.792	0.3894	−0.0441	HE601673
Monomorphic											
Mss 15	$(GA)_9$*	F: ACGCAACCAAAAGAAAGTGC R: TCCGTTCAAAGAGACACGAA	52.4								HE601649

NAF values above 0.100 are showed in bold.
*Imperfect microsatellite.

© 2012 Blackwell Publishing Ltd

Table 2 Expressed sequence tag (EST)-SSRs isolated from Senegalese sole with repeat motif, primer sequences, PCR annealing temperature (T), number of alleles (NA) and allele range in pair bases (pb), observed heterozygosity (H_O), expected heterozygosity (H_E), exact-probability test (*P*-value) and null-allele frequency (NAF)

Locus	Repeat motif in cloned sequences	Primer sequences (5′–3′)	T (°C)	NA	Allele range (pb)	H_O	H_E	Polymorphism information content	*P*-value	NAF	Accession number
Polymorphic											
EST-1I17	$(TG)_{20}(CG)_8$	F: AAGAATAGCTGCCCAAACC R: TGTTTTCATTTAGTTGTATATGTGAGA	55	13	83–125	0.8462	0.9087	0.881	0.0216	+0.0231	HE600105
EST-16C12	$(CA)_{41}(CG)_8$	F: CATAAATAATCGGGGGATTCT R: CCGACTCATGTTGCTTTGTT	59.5	14	74–258	0.3913	0.8831	0.851	0.0000	**+0.3811**	HE600106
EST-2H15	$(TG)_{23}$	F: ACCAAAGTAGCGCAGATTCC R: CTTCATCAGCAGCCAAACTG	59.5	12	103–157	0.7917	0.7624	0.727	0.7930	−0.0253	HE600107
EST-1N07	$(GT)_9$*	F: TGCACATCAGTGAGTTAATATT R: T TGTGATGGCGTGAAAAGTTC	55	3	125–133	0.1053	0.1991	0.185	0.0534	**+0.2879**	HE600108
EST-3H07	$(TG)_{15}$*	F: CCCAATTACAATAGTGGCCTGT R: CCTTCAATGCTTCAGCTGTCT	58	8	88–107	0.7407	0.6778	0.637	0.8111	−0.0550	HE600109
EST-2G14	$(TG)_{19}$*	F: ACCAAAGTAGCGCAGATTCC R: CGAGAGCTTAACACCACAGC	59.5	11	79–132	0.8000	0.7423	0.699	0.3611	−0.511	HE600110
EST-3A04	$(CGT)_{11}(GT)_{26}$*	F: AAATACGAGGGTCGTCACAAA R: GGATTACAGCAAAGTGGAGTGA	62	15	161–199	0.9231	0.8401	0.811	0.9072	−0.0749	HE600111
EST-6A20	$(TAAA)_5(CAAA)_5$*	F: TGGGAGAGGTAGAAGCATGG R: CTCAAGTGGACATGCCAAGA	60.5	4	82–94	0.4000	0.5853	0.524	0.0103	**+0.2073**	HE600112
EST-22E12	$(CT)_{18}$*	F: GCTGGAAGCTGCACACG R: GACGAAAGCGTTTTGTCAG	63	4	103–111	0.5200	0.6433	0.573	0.1101	**+0.1032**	HE600113
EST-1P20	$(GT)_{23}$*	F: GAAAGGACGGCGGTTGCAC R: CAAGAATGTATATTGGATGAAAGACA	57	11	110–133	0.9167	0.8794	0.846	0.9214	−0.0314	HE600114
EST-27G19	$(TA)_{16}(CA)_{27}$*	F: CACACTGTCAGGAACTACAGAGG R: GGAGAAATGAAATTGGATTTTAA	56.5	10	138–173	0.9200	0.8237	0.788	0.7368	−0.0699	HE600115
EST-7D10	$(CT)_{12}$*	F: CCTCTTTAAATATGATTCCTTTACATG R: AGCATCAGTGACAATGTTGTTTC	57.5	6	100–117	0.5556	0.6660	0.609	0.8318	+0.0534	HE600116
EST-5	$(TGGACA)_7$	F: ATGCCAATAAACCTGGCACT R:GAACCGATCCCAACTGTGTT	55	3	162–174	0.583	0.550	0.480	0.1621	−0.0689	HE600077
EST-8	$(TG)_{18}$*	F: TCATGGTAAGCAAGGTGCAA R: ATATGCACCAGATGCTGGAC	57	8	154–168	0.840	0.789	0.743	0.6645	−0.0410	HE600073
EST-10	$(GGA)_{13}$*	F: CGGAGTCAAACATTCACTCAAA R: GTGGTCGACGGAATCAAACT	55	6	186–207	0.800	0.773	0.717	0.9808	−0.0319	HE600075
EST-11	$(GT)_9$	F: TCGAGTGGACAACACTACGC R: GGGTGAAACTTCCCCATTCT	55	3	149–153	0.640	0.636	0.552	0.4813	−0.0014	HE600076
EST-12	$(CA)_{11}$	F: AAGATAACCCCCGTGTTGTG R: GACCGTTAAAACTCCCCACAT	55	6	182–196	0.720	0.807	0.761	0.0323	+0.0536	HE600078
EST-13	$(GGA)_8$	F: ATCTGACCTTCCCCTCCATC R: TTTTTCCAACTGGTGCTTTTT	55	6	144–159	0.920	0.777	0.726	0.2790	−0.1023	HE600079
EST-14	$(CA)_8$	F: AAATAATACAGAAAGATGCCTTCAA R: AGCAGCCTGAAGCAGGACTA	55	4	146–152	0.480	0.453	0.405	0.2304	−0.0192	HE600080

© 2012 Blackwell Publishing Ltd

Table 2 (Continued)

Locus	Repeat motif in cloned sequences	Primer sequences (5′–3′)	T (°C)	NA	Allele range (pb)	H_O	H_E	Polymorphism information content	*P*-value	NAF	Accession number
EST-15	$(ACTC)_8$	F: TGCAAAAAGTTGAGGCTCATAA R: TCCTGGACTGTTTTTCACTTTG	55	3	208–216	0.409	0.382	0.337	0.7345	−0.0296	HE600081
EST-16	$(TG)_{14}$	F: CAGAGGAACCGTCGACACTC R: TTGTGTTGCAGTTTCTGTCG	50	11	90–118	0.680	0.884	0.853	0.0069	**+0.1210**	HE600082
EST-17	$(GT)_9$	F: CCAGAGGAACAGTCGACACTC R: CAGTTGTGTTGCAGCTTCTG	50	12	93–121	0.720	0.873	0.841	0.0946	+0.0895	HE600083
EST-18	$(CA)_{10}$	F: AAACTCTGCCGTGTGATGTG R: TCCACTCGTGCAAGCTAAGA	55	8	195–207	0.680	0.830	0.793	0.0000	+0.0874	HE600083
EST-22	$(TG)_{13}$	F: CCCATCGTTGTGGTTCTTCT R: CTCCTGTATTCCCAGGTCCA	55	6	142–162	0.783	0.761	0.708	0.3257	−0.0289	HE600085
EST-23	$(CA)_8$	F: CTCAGCCTCTCCTTCATTCC R: CAGTTTGCGCTGACAACATAA	55	6	124–134	0.773	0.756	0.703	0.7583	−0.0292	HE600086
EST-26	$(CA)_9$	F: TCCACTTGTCTTTATTGAACACATT R: AAATCAAAGCAGGGCATCAT	55	4	152–158	0.760	0.701	0.632	0.4702	−0.0446	HE600088
EST-32	$(CA)_{15}$	F: TCTGAAAACCTGAGGTGACG R: TTTCTCCCGTGCTAAGATGG	55	11	227–247	0.920	0.860	0.827	0.1270	−0.0487	HE600090
EST-33	$(GAA)_9$	F: GCTTCAGCAAACAACAGCAA R: TGCTTTTCATGTTTCAAACTCC	55	6	142–157	1.000	0.744	0.691	0.0003	−0.1820	HE600091
EST-36	$(TA)_8$	F: GTGTTTAGAAAAACACAAACTTATCAA R: CCTGTCAGTGTTTGTTTGAAGG	55	3	98–102	0.450	0.550	0.477	0.1573	+0.0658	HE600092
EST-37	$(TTA)_8$	F: TGGTCGTTTTCATTCCGATA R: AACAATGGCTAGCCTGAGATG	55	2	194–197	0.417	0.454	0.346	1.0000	+0.0323	HE600093
EST-43	$(GA)_8$	F: CAAAAACAAAATCAACAGTGCAA R: TTCCGTCCATCTTTTCACCT	55	5	172–182	0.360	0.410	0.377	0.0464	**+0.1146**	HE600094
EST-47	$(GT)_{11}$	F: TGGGAGAGGTCAGGAGTACG R: AGACTTCACACCGGGATCAG	55	11	461–497	0.958	0.914	0.886	0.0786	−0.0349	HE600096
EST-48	$(GT)_{13}$	F: AGTCTCTGCCCAGCCAAAT R: AGCTGTAGGCCCCAACC	55	5	156–178	0.760	0.724	0.662	0.3541	−0.0401	HE600097
EST-50	$(CGT)_9$	F: CGTCTTTCCTGTGGTGGTTT R: CTCCTACATCCCCCTTTTCC	55	2	153–156	0.200	0.497	0.369	0.0036	**+0.4179**	HE600098
EST-60	$(GT)_{15}$	F: AAGCAAACATTCATCCGTCA R: ATTGATTTACCCAACAGCGTCT	55	7	134–148	0.800	0.819	0.776	0.5057	−0.0036	HE600102
EST-62	$(CAT)_{20}(TCG)_9$*	F: CAGCTGCTTGAAGTCCATGT R: GGCAAAGTCAACACTGAA	52	5	205–217	0.750	0.781	0.724	0.9282	+0.0147	HE600103
EST-65	$(CA)_{34}$	F: AGGGTCAGGCTGCTTACTTG R: TCATTCTGTGCTTCTGTGCTG	53	15	181–309	0.400	0.917	0.890	0.0000	**+0.3867**	HE600074
Monomorphic											
EST-8D21	$(TACA)_4$	F: GGACCTTAATGCTGGACCTT R: CAACATAATGCTGCCCACTG	62								HE600117

© 2012 Blackwell Publishing Ltd

Table 2 (Continued)

Locus	Repeat motif in cloned sequences	Primer sequences (5′–3′)	T (°C)	NA	Allele range (pb)	H_O	H_E	Polymorphism information content	*P*-value	NAF	Accession number
EST-2M07	$(CTT)_8(GT)_3$*	F: CATCTATCGAGCTGTT R: AGTCACCGCAGCCCC	61								HE600118
EST-11G02	$(TAAA)_8$*	F: CCTCGTGGAAGAGGTCTTGA R: GAGGAAGAGCTCAAAACAAA	61								HE600119
EST-21P06	$(CTT)_{13}$*	F: TGTCTGCAGATGAGCCAGAT R: GAGTGACGATGAGGGAGAGG	53								HE600120
EST-19	$(GGA)_{11}$	F: TTGGGAGGGGGTTCACAG R: AGCTGCTCAAAGACGGAGAC	58								HE600084
EST-25	$(TTTA)_4$	F: CGAATGACGCCAAACTACAC R: CTCGCATCTAGACACGGAGA	55								HE600087
EST-30	$(GA)_{11}$*	F: TTTCTGAGCTGCAAATAATACATTTC R: TGTGCCCATCCACATGTTAG	55								HE600089
EST-45	$(GA)_{15}$	F: CGGCACGAGGAGTTTAAAAG R: TGTGGAGAAAAGGAGGGTTG	60								HE600095
EST-46	$(GA)_{15}$	F: GGTACGTCGTCATCCTCCTC R: AGCGCCACCTATAGGAATGA	52								HE600095
EST-53	$(ATGA)_4$	F:TGAGATGTGAACGAGACAAAGG R: TGGACAGATGTGTCTCCACCT	55								HE600099
EST-56	$(CTG)_{10}$	F: GGAATTTAAAGAACAGAAAAGAGACA R: GATCGAGGGTCACACCATTT	50								HE600100
EST-58	$(GA)_9$	F: GTACGGATCAGCACTGTGGA R: TGCATTCCCTGCATTTTG	55								HE600101
EST-66	$(CA)_{11}$	F: GCCAGGACTGATATCGCATT R: GTGTTGGTTGCCTCCTGTTT	57								HE600104

NAF values above 0.100 are showed in bold.
*Imperfect microsatellite.

© 2012 Blackwell Publishing Ltd

constituted mainly by dinucleotide repetitions (62.5% of the anonymous ones and 75.7% of the EST-SSRs belong to this category); however, oddly, trinucleotide markers were more frequent in anonymous SSRs (28%, against 19%). Both sets showed fairly similar values for several characteristics. Thus, within the EST-SSR set, the number of alleles ranged from 2 to 15 (7.27 ± 3.86 alleles per locus), whereas the observed and expected heterozygosities varied from 0.105 to 1.000 (0.667 ± 0.22 per locus) and from 0.199 to 0.917 (0.709 ± 0.17), respectively. The genomic SSR set displayed rather similar results for these parameters, varying between 2–19 (8.96 ± 4.81), 0.052–0.954 (0.625 ± 0.22) and 0.051–0.936 (0.680 ± 0.21), respectively. PIC and average probability test values per locus also showed no significant differences between the two sets, being 0.659 ± 0.18 and 0.377 ± 0.35 for the EST-SSR set, and of 0.632 ± 0.22 and 0.411 ± 0.28 for the genomic SSR set, respectively. On the other hand, despite that the EST sequences appear to be more conserved than anonymous ones, the number of markers that showed significant heterozygote deficiency (NAF > 0.1) was quite similar for both sets (Tables 1 and 2), possibly due to the presence of null alleles. Results from linkage disequilibrium analyses among the polymorphic maker pairs have been included as Table S1. Genotypic disequilibria were detected in 52 of 2300 pairwise comparisons ($P < 0.05$), that is, in the 2.26% of the cases, a proportion lower than the expected 5% (115 significant associations).

The BLASTX homology search revealed strong similarity between annotated genes and some microsatellite-containing sequences (Table 3). As expected, those

Table 3 Annotation of monomorphic and polymorphic Senegalese sole expressed sequence tag (EST) and anonymous microsatellites, using BLASTX comparisons

Microsatellites	Closest identity	Accession no.	Identity (%)	*E*-value
EST-derived microsatellites				
EST-1I17	*Danio rerio* hypothetical protein	NP_001116761.1	60	4e-17
EST-2H15	*Salmo salar* casein kinase II subunit alpha	NP_001133529.1	88	3e-42
EST-3H07	*S. salar* interferon-related developmental regulator 2	NP_001167307.1	83	2e-36
EST-2G14	*S. salar* casein kinase II subunit alpha	NP_001133529.1	88	3e-42
EST-3A04	*Osmerus mordax* PHS1	ACO09320.1	94	6e-70
EST-22E12	*Tetraodon nigroviridis* unnamed protein	CAG11505.1	86	9e-42
EST-27G19	*D. rerio* usp7 protein	AAH94294.1	100	1e-04
EST-2M07*	*Oreochromis niloticus* protein FAM193A-like	XP_003449784.1	100	1e-31
EST-5	*O. niloticus* C1q-like protein	XP_003438219.1	49	3e-09
EST-8	*S. salar* tropomyosin-1 alpha chain	ACI34179.1	95	2e-15
EST-10	*O. mordax* creatine kinase, testis isozyme	ACO08899.1	88	5e-34
EST-11	*Esox lucius* trafficking protein particle complex subunit 2-like protein	ACO13683.1	95	9e-82
EST-13	*Pennahia argentata* myosin light chain 1	BAA95129.1	91	8e-84
EST-15	*D. rerio* cornifelin homolog	NP_001004663.1	76	8e-18
EST-16	*Scophthalmus maximus* 40S ribosomal protein S27	ABJ98653.1	98	3e-48
EST-19*	*O. niloticus* troponin C	XP_003457006.1	96	4e-91
EST-22	*T. nigroviridis* unnamed protein	CAF96756.1	97	9e-77
EST-23	*T. nigroviridis* unnamed protein	CAF96756.1	95	5e-115
EST-26	*Epinephelus coioides* 14 kDa apolipoprotein	ACM41842.1	64	5e-46
EST-32	*E. coioides* malate dehydrogenase 1b	ACL98112.1	92	1e-89
EST-33	*Cyprinus carpio* S31 protein	CAA76640.1	96	8e-42
EST-45*	*O. niloticus* transcription factor BTF3	XP_003448501.1	90	1e-04
EST-46*				
EST-47	*T. nigroviridis* unnamed protein	CAG14833.1	86	2e-87
EST-50	*O. mordax* PHS1	ACO09320.1	91	2e-97
ESST-56*	*Perca flavescen*s mitochondrial NADH dehydrogenase flavoprotein 1	ADX97063.1	94	1e-148
EST-58*	*O. niloticus* la-related protein 1B-like	XP_003447897.1	68	5e-22
EST-60	*T. nigroviridis* unnamed protein	CAG13169.1	89	4e-19
EST-62	*Solea senegalensis* calsequestrin 2	BAG49513.1	100	1e-107
Anonymous microsatellites				
Mss24	*D. rerio* proton-coupled amino acid transporter 1	XP_687732.1	86	5e-07
Mss32	*D. rerio* semaphorin	NP_001186294.1	94	4e-13
Mss44	*T. nigroviridis* unnamed protein	CAG07831.1	95	1e-14

*Monomorphic SSR.

© 2012 Blackwell Publishing Ltd

containing EST-SSR (including both monomorphic and polymorphic) were annotated in a much greater proportion (58%) than the anonymous ones (9%).

According to the Gene Ontology annotation, most annotated sequences are related with the following cellular components: membrane (16%), endoplasmic reticulum (12%), nucleus (8%), ribosome (8%) and cytoplasm (4%), showing a distribution similar to those found in turbot (Navajas-Pérez *et al.* 2012).

The annotated sequences were also analysed to determine the relative position of the microsatellite; thus, most SSR (96.5%) proved to be located within untranslated regions (UTR), mainly within 5′-UTR (58.62%). Microsatellites located in 5′ and 3′ UTR could affect gene expression through their influence on the stability of transcribed products and on the translation level from the target RNA, as described in European eel (*Anguilla anguilla*), where the repeats of an SSR located at the 3′-UTR of the *TSHB* gene are able to form unusual helix structures within double-stranded DNA that promote genetic instability (Pradet-Balade *et al.* 1998; Chistiakov *et al.* 2006). In the present work, only one microsatellite (EST-5) was located within the translated region; it consists of a hexanucleotide motif (TGGACA that allows unbroken reading of the frame) placed within the gene that supposedly encodes the C1q protein, a subunit of the C1 enzyme complex that activates the serum complement system. Further studies will determine whether there is relation between the different alleles present and the level of production of this glycoprotein.

Additionally, all EST and the annotated anonymous sequences (50 and 3, respectively) were BLASTed against Atlantic halibut, Japanese flounder and turbot EST databases to evaluate the degree of conservation of these SSRs in other flatfish species. All the anonymous sequences analysed gave negative results; however, 23 EST-SSR (46%) showed significant matches with any of the EST databases explored (Table 4). Of these, only one (EST-10) matched with all three species; four (17.4%) matched with two species, and the remaining 18 (78.3%) matched with only one species. Most matches were recorded with the turbot database (14, against 8 for each of the other databases). Alignment of these sequences showed that although the SSR-flanking region was certainly conserved (to a variable degree), the microsatellite itself was not always present. Thus, 71% of the SSR that hit with the turbot EST database lacked the respective repetitive motif. On the contrary, most of the hits with the halibut and flounder databases indeed included it (62.5% and 75%, respectively). In this sense, it bears noting that the presence or absence of the SSR motif was variable even within a given EST-SSR. Thus, the EST-12, which matched with flounder and turbot EST databases, included the SSR motif with the first species, but not with the second. In other cases, no variation appeared: EST-10, EST-19 and EST-32 included the SSR motif in all the species where they were found, whereas in the same way, EST-22 and EST-23 lacked it in the two species where they were detected. On the other hand, although some of the Senegalese sole EST-SSR described here appear to be conserved in other flatfish species (e.g. EST10, being present in all three species, and EST19, being present in halibut and flounder), the SSR-flanking sequences showed minor differences that could prevent the use of the designed primers with these related species (Fig. 1); in fact, low cross-species amplification has previously been reported for the order Pleuronectiformes (Bouza *et al.* 2002; Castro *et al.* 2006). Despite this, the *in silico* information reported here would be of interest to develop new SSR, equivalent to those found in Senegalese sole, in these flatfish species, allowing comparative genomic studies.

Table 4 Senegalese sole expressed sequence tag (EST)-SSR-containing sequences that showed significant matches with EST databases sequences from Atlantic halibut, Japanese flounder and turbot

	Atlantic halibut	Japanese flounder	Turbot
Monomorphic SSR	EST-19, EST-21P06, **EST-58**, EST-66	EST-19, EST-2M07	**EST-56**
Polymorphic SSR	EST-10, **EST-22**, **EST-23**, EST-62	EST-10, EST-12, EST-32, EST-48, **EST-2H15**, **EST-2G14**	**EST-8**, EST-10, **EST-11**, **EST-12**, EST-14, **EST-16**, **EST-22**, **EST-23**, EST-26, EST-32, **EST-33**, **EST-60**, **EST-27G19**

Cases in which the SSR-flanking sequences were conserved but the SSR itself was absent are indicated in bold.

In conclusion, here we describe two new sets of Senegalese sole SSRs, including the first found in EST sequences. EST-SSR markers are considered to have many advantages over genomic SSR markers, for example a higher proportion of high-quality markers or a higher transferability among related species (Varshney *et al.* 2005). Our EST sequences certainly provided a higher proportion of quality markers (74% of the tested SSR were useful, against 64% in anonymous ones), all the same differences was not significant. On the other hand, given the discrepancies observed between the SSR-flanking sequences of the four flatfish species analysed, it is doubtful that cross-species assays using the proposed SSR primers will succeed. However, other primers targeting these conserved SSR could allow the performance of

© 2012 Blackwell Publishing Ltd

Fig. 1 Partial alignment of the Senegalese sole sequence expressed sequence tag (EST)-10 with sequences taken from EST databases of Japanese flounder (FE043204), Atlantic halibut (DN794634) and turbot (FE945246). Microsatellite motif (GGA) is shown in bold. Location of the corresponding primers is shown in grey.

comparative studies. On the other hand, EST-SSR offers other advantages compared with genomic ones, because they can be more rapidly developed and can be frequently associated with genes of known functions (Chistiakov *et al*. 2006). In this sense, EST-SSRs are also of interest because they offer the opportunity to determine the effect of repeat polymorphism on gene expression. The new polymorphic microsatellite sets described here, together with others previously described, will enable studies of genetic diversity, genome mapping and molecular-assisted breeding of Senegalese sole, a valuable fish in the aquaculture industry.

Acknowledgements

This work has been supported by the Spanish Ministry of Science and Innovation (AGL2009-11872) and the Consolider-Ingenio AQUAGENOMICS Project (CSD2007-00002).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Bouza C, Presa P, Castro J, Sánchez L, Martínez P (2002) Allozyme and microsatellite diversity in natural and domestic populations of turbot (*Scophthalmus maximus*) in comparison with other Pleuronectiformes. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 1460–1473.

Bouza C, Hermida M, Pardo BG *et al.* (2007) A microsatellite genetic map of the turbot (*Scophthalmus maximus*). *Genetics*, **177**, 2457–2467.

Casas L, Sánchez L, Orban L (2009) Novel microsatellites from the European plaice (*Pleuronectes platessa*)-identification by data mining and cross-species amplification in other flatfishes. *Conservation Genetics*, **10**, 1565–1568.

Castro J, Pino A, Hermida M *et al.* (2006) A microsatellite marker tool for parentage analysis in Senegal sole (*Solea senegalensis*): genotyping errors, null alleles and conformance to theoretical assumptions. *Aquaculture*, **261**, 1194–1203.

Chen S-L, Shao C-W, Xu G-B, Liao X-L, Tian Y-S (2008) Development of 15 novel dinucleotide microsatellite markers in the Senegalese sole *Solea senegalensis*. *Fisheries Science*, **74**, 1357–1359.

Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*, **255**, 1–29.

Coimbra MRM, Kobayashi K, Koretsugu S *et al.* (2003) A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture*, **220**, 203–218.

De la Herrán R, Robles F, Navas JI *et al.* (2008) A highly accurate, single PCR reaction for parentage assignment in Senegal sole based on eight informative microsatellite loci. *Aquaculture Research*, **39**, 1169–1174.

Funes V, Zuasti E, Catanese G, Infante C, Manchado M (2004) Isolation and characterization of ten microsatellite loci for Senegal sole (*Solea senegalensis* Kaup). *Molecular Ecology Notes*, **4**, 339–341.

Imsland AK, Foss A, Conceição LEC *et al.* (2003) A review of the culture potential of *Solea solea* and *S. senegalensis*. *Reviews in Fish Biology and Fisheries*, **13**, 379–407.

Kalinowski ST, Taper ML, Marshall TC (2006) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **6**, 974–979.

Kang JH, Kim WJ, Lee WJ (2008) Genetic linkage map of olive flounder, *Paralichthys olivaceus*. *International Journal of Biological Sciences*, **4**, 143–149.

Liu YG, Zheng MG, Liu LX *et al.* (2006) Five new microsatellite loci for Oliver flounder (*Paralichths olivaceus*) from an expressed sequence tag (EST) library and cross-species amplification. *Molecular Ecology Notes*, **6**, 371–373.

Moriñigo MA, Zorrilla I, Chabrillon M, Arijo S, Diaz R, Balebona MC (2001) Isolation and characterization of *Vibrio harveyi* strains causing infection in sole (*Solea senegalensis*). 10th International Conference of the European Association of Fish Pathologists. Dublin.

Navajas-Pérez R, Robles F, Molina-Luzón MJ *et al.* (2012) Exploitation of a turbot (*Scophthalmus maximus* L.) immune-related expressed sequence tag (EST) database for microsatellite screening and validation. *Molecular Ecology Resources*, Doi: 10.1111/j.1755-0998.2012-03126.x..

Pardo BG, Fernández C, Millán A *et al.* (2008) Expressed sequence tags (ESTs) from immune tissues of turbot (*Scophthalmus maximus*) challenged with pathogens. *BMC Veterinary Research*, **4**, 37.

© 2012 Blackwell Publishing Ltd

Porta J, Alvarez C (2004) Development and characterization of microsatellites from Senegal sole (*Solea senegalensis*). *Molecular Ecology*, **4**, 277–279.

Pradet-Balade B, Salmon C, Hardy A, Querat B (1998) Heterogeneity of eel thyrotropin beta mRNAs is due to a minisatellite in the 3′ untranslated region of the gene. *Gene*, **215**, 251–257.

Raymond M, Rousset F (1995a) An exact test for population differentiation. *Evolution*, **49**, 1283–1286.

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

Reid DP, Smith C-A, Rommens M, Blanchard B, Martin-Robichaud D, Reith M (2007) A genetic linkage map of Atlantic Halibut (*Hippoglossus hipoglossus* L.). *Genetics*, **107**, 753–774.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NewYork, NY, USA.

Sekino M, Saitoh K, Yamada T, Kumagai A, Hara M, Yamashita Y (2003) Microsatellite-based pedigree tracing in a Japanese flounder *Paralichthys olivaceus* hatchery strain: implications for hatchery management related to stock enhancement program. *Aquaculture*, **221**, 255–263.

Sha ZX, Luo XH, Liao XL, Wang SL, Wang QL, Chen SL (2011) Development and characterization of 60 novel EST-SSR markers in half-smooth tongue sole *Cynoglossus semilaevis*. *Journal of Fish Biology*, **78**, 322–331.

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*, **23**, 48–55.

Vera M, Alvarez-Dios JA, Millán A *et al*. (2011) Validation of single nucleotide polymorphism (SNP) markers from an immune expressed sequence tag (EST) turbot, *Scophthalmus maximus*, database. *Aquaculture*, **313**, 31–41.

M.J.M.L., J.R.L. and F.R. carried out the technical work to develop the libraries and to validate different sets of microsatellites in the laboratory. M.J.M.L., J.R.L. and R.H. mined EST sequences from EST database and anonymous libraries to look for microsatellites and designed the primers. R.N.P. was the responsible of developing the bioinformatics tools for microsatellites analyses. R.N.P., R.H. and CR coordinated the work and wrote the manuscript.

Data accessibility

DNA Sequences: GeneBank accessions HE600073-HE600122, and HE601639-HE601673; see Tables 1 and 2 for details. Sequence alignments are provided as Table S1.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. Linkage disequilibrium analyses among the polymorphic maker pairs.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2012 Blackwell Publishing Ltd