Characterization of 27 novel gene-associated SSR markers in Indian catfish, Clarias batrachus (Linnaeus, 1758) and their...

Article *in* Molecular Ecology Resources · December 2012

CITATION
1

READS
95

7 authors, including:

Vindhya Mohindra
National Bureau of Fish Genetic Resources, Luc...
137 PUBLICATIONS **760** CITATIONS
SEE PROFILE

Ruchi Patangia
University of Central Missouri
4 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Ratnesh Tripathi
National Bureau of Fish Genetic Resources
16 PUBLICATIONS **65** CITATIONS
SEE PROFILE

Kuldeep Lal
Network of Aquaculture Centers in Asia-Pacific
153 PUBLICATIONS **646** CITATIONS
SEE PROFILE

Characterization of 27 novel gene-associated SSR markers in Indian catfish, *Clarias batrachus* (Linnaeus, 1758) and their application in genetic diversity analysis

Authors: Vindhya Mohindra, Akanksha Singh, Ruchi Patangia, Ratnesh K. Tripathi, Rajeev K. Singh, Rama Shankar Sah and Kuldeep K. Lal

National Bureau of Fish Genetic Resources (ICAR),
Canal Ring Road, P. O. Dilkusha, Lucknow 226002, UP, India

Corresponding author: **Vindhya Mohindra**
National Bureau of Fish Genetic Resources (ICAR),
Canal Ring Road, P. O. Dilkusha, Lucknow 226002, UP, India

E mail address: vindhyamohindra@gmail.com

Contact: Tel: +91 522 2442440, 2442441 Fax: +91 522 2442403

Keywords: *Clarias batrachus*; Expressed sequence tags; EST-SSR, polymorphic loci; heterozygosity

Running title : EST-SSR markers in *Clarias batrachus*

ABSTRACT

27 polymorphic EST-SSRs were used to analyze two natural populations of *Clarias batrachus* for population variability parameters. Allele numbers ranged from 2 to 12; and observed and expected heterozygosity values from 0.3658 and 0.6850; and 0.5668 and 0.5828, respectively. Fifteen loci showed no significant departure from HWE ($p<0.05$). Linkage disequilibrium test was significant for three pairs of loci. Two populations studied were found to be significantly divergent. Cross-species amplification for 7 loci was successful in two related species. These SSRs markers may be useful for the assessment of genetic variations and population differentiation studies in *C. batrachus*.

Keywords: Indian catfish; *Clarias batrachus*; Expressed sequence tags; polymorphic loci; heterozygosity

Clarias batrachus (Linnaeus, 1758), (*C. magur* Hamilton (1822), an Indian catfish species, is endemic to the Indian subcontinent (Pouyaud et al. 2009). This hardy, omnivorous and air breathing fish inhibits wetlands, swamps, rivers ponds and tanks, and is well adapted to adverse ecological conditions. It is popular owing to its taste, medicinal value and high market price (Chonder 1999). The species is declining in its abundance primarily due to drying up of wetlands and is considered as an endangered species (Vishwanath 2010). Therefore, conservation strategies including propagation-assisted rehabilitation of natural populations are being considered as a priority. Population structure information obtained through polymorphic markers with consistent scorable alleles in *C. batrachus* will be useful in planning conservation strategies.

Microsatellites or simple sequence repeats (SSR) are widely used in population genetic analysis because of their high polymorphism, abundance and co-dominance (Liu & Cordes 2004). Expressed sequences provide an alternative strategy for identification of gene-associated microsatellite (EST-SSR) markers, which are efficient tools for discovering functional polymorphisms, especially in non-model species (Vasemagi et al. 2005). Moreover, EST-SSRs generally have fewer null alleles, greater cross species amplification and less allelic variability than genomic SSRs (Pashley et al. 2006). Despite high commercial and evolutionary significance, development of the genomic resources including gene-associated markers in clarid catfishes is very limited, except few (9) polymorphic microsatellites (Volckaert et al. 1999).

In this study, 1937 EST sequences (Mohindra et al. 2012) were screened for the presence of SSRs. Of the identified 221 microsatellite containing unique ESTs, 78 ESTs were selected for primer designing on the basis of representation of all types of repeats with considerable repeat number, so that there is more possibility of loci being polymorphic. To amplify the repeat regions, primers were designed using the web based tool Primer3 (http://primer3.sourceforge.net/) (Rozen & Skaletsky 2000) to amplify a PCR product of approximately 120-150 bp, with an optimum T_a of 55°C and a minimum GC content of 30-70%. PCR amplification was performed in a 25-µL reaction using PTC200 thermocycler (MJ Research) as follows: 5 min denaturation at 95 °C; 25 cycles : 30 s at 95 °C; 30 s at 50°C and 60 s at 72 °C, with a final extension of 10 min at 72 °C. The PCR mix

consisted of 50 ng of genomic DNA, 1× PCR buffer (10 mm Tris-HCl, pH 9.0; 50 mm KCl; 0.01% gelatin), 5 pM of each primer, 1.5 mm $MgCl_2$, 200 μM dNTPs and 1.5 U Taq DNA Polymerase (Genei, India). Amplified products were resolved on non-denaturing polyacrylamide gel followed by silver staining (silver staining kit, Pharmacia Biotech, USA). Primer pairs for the identified 78 loci were initially tested on 8 genomic DNA samples of *C. batrachus*. To evaluate the suitability of polymorphic loci for genetic diversity analysis, genomic DNA of 48 specimens collected from two geographically distinct locations, Lucknow (26º 50' N; 80º 55' E) n=23 and Varanasi (25º 16' N; 82º 57' E) n=25, Uttar Pradesh, India were genotyped with the identified polymorphic microsatellite loci. Genomic DNA (n=2 to 5) of *Clarias gariepinus*, *Clarias dussumieri* and *Heteropneustes fossilis* (Table 3) were tested for cross priming of 12 EST-SSRs, to identify the utility of EST-SSR loci in finding the homologus loci in related species.

Individual genotype data at each of the 27 polymorphic microsatellite loci were analysed using the software Genetix Version 4.05 (Belkhir et al. 2004) to determine parameters of genetic variation and differentiation. Tests for linkage disequilibrium (LD), conformity to Hardy–Weinberg equilibrium and genotypic differentiation were performed with software Genepop Version 3.4 (Raymond & Rousset 1995). Presence of null alleles was assessed with the software Micro-Checker (Van Oosterhout et al. 2004).

Of the 78 unique EST-SSRs primer pairs tested, all the primer pairs, except three, successfully amplified loci in expected size range. Of these, 27 (36%, table 1) were found to be polymorphic in the two isolated populations screened in this study. The number of alleles in these loci ranged from 2-12 per locus and mean number of alleles per locus was 4.4815 (Lucknow) and 4.5556 (Varanasi). Expected heterozygosities values were 0.5668 (Lucknow) and 0.5828 (Varanasi) and observed heterozygosities were 0.3658 (Lucknow) and 0.6850 (Varanasi). There was evidence of no significant deviation from Hardy–Weinberg expectations (Table 1) at fifteen loci; CbSpn0449, cb0781, CbSpn1100, CbSpn1115, CbSpn1191, CbSpn1837, CbSpn2339, CbSpn2440, CbSpn25, CbSpn2524, CbSpn2670, CbSpn3499, CbSpn4566, CbSpn956 and CbSpncontig6 in the two populations, after the probability level ($p<0.0038$) was adjusted for Sequential Bonferroni (Lessios 1992). Assessment with the software Micro-Checker using Bonferroni confidence interval

revealed the possible signs of null alleles at 3 loci, that could be responsible for the observed excess of homozygotes (+Fis). Significant genetic heterogeneity ($p < 0.05$) between the two populations was evident at ten loci, CbSpn2670, CbSpn1817, CbSpn268, CbSpn1115, CbSpn1191, CbSpn2440, CbSpn4654, CbSpn4566, CbSpncontig6 and CbSpncontig12, and genotypic differentiation (FST, Weir & Cockerham, 1984) for each population pair across all loci was 0.08626 ($p<0.001$). The results of Bayesian clustering indicated that all the individuals could be assigned to two distinct groups without any indication of admixture population. Test for linkage disequilibrium was found to be statistically significant ($p<0.05$) for three pairs of EST-SSR loci (CbSpn0449 & CbSpn2339; CbSpn2339 & CbSpn4620; CbSpn4620 & CbSpncontig6) when tested per population. Thus, 13 EST-SSR loci were found to be in HWE and unlinked. The genes, in which these SSRs are found, are located in different linkage groups in Zebrafish (Flicek et al. 2012) and Kucuktas et al. (2009) established putative conserved syntenies between catfish and the model fish species through microsatellite and SNP markers on genetic maps. Thus, the possible reason of linkage observed may be due to small sample size. Significant genetic heterogeneity and the Fst values as well as Beysian clustering indicated the significant genetic heterogeneity for the samples from Lucknow and Varanasi. Thus, using EST-SSRs two genetically divergent populations of *C. batrachus* could be observed in this study.

Success rate of cross species amplification of a given primer in three related species tested ranged from 0 to 50% (C. *dussumieri* upto 50%, *C. gariepinus* 0%, and *H. fossilis* 2.5%) (Table 2). Out of the 12 loci tested, 7 successfully amplified in *C. dussumeri* (6 loci) and *H. fossilis* (3 loci). Two loci, CbSpn268 and CbSpn143, amplified both in C. *dussumieri* and *H. fossilis*. However, none of the loci cross amplified in *C. gariepinus* samples.

The present results clearly demonstrated that these polymorphic EST-SSRs proved to be useful for the assessment of genetic variations and population differentiation studies in *C. batrachus*.

Acknowledgements

This study was conducted under Department of Biotechnology , GOI funded and Institute's NBFGR-ICAR G/FC/01/09 projects. Excellent technical assistance

provided by Mr. Rajesh Kumar and Mr. Sree Ram is duly acknowledged. The authors are thankful to Dr. A. Gopalakrishnan for providing Clarias dussumieri samples.

References

Belkhir K., Borsa P., Goudet J., Chikhi L. & Bonhomme F. (2004) genetix 405 logiciel sous Windows pour la génétique des populations *Laboratoire Génome Populations Interactions CNRS UMR 5171* Université de Montpellier II Montpellier France. http://wwwgenetixUniversitymontp2fr/genetix/genetix htm .

Chonder S. L. (1999) Biology of Finfishes and Shellfishes. SCSC Publishers Howrah.

Flicek P., Amode M.R. & Barrell D. et al. Ensembl (2012) Nucleic Acids Res, 40 (Database issue): D84-90. Database URL: http://www.ensembl.org/index.html.

Kucuktas H., Wang S., Li P., He C., Xu P., Sha Z., Liu H., Jiang Y., Baoprasertkul P., Somridhivej B., Wang Y., Abernathy J., Guo X., Liu L., Muir W. & Liu Z. (2009) Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. Genetics 181 : 41649-1660.

Lessios H.A. (1992) Testing electrophoretic data for agreement with Hardy–Weinberg expectations. Mar Biol 112 : 517–523.

Liu Z.J. & Cordes J.F. (2004) DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238: 1-37.

Mohindra V., Singh A., Barman A. S., Tripathi R., Sood N. & Lal K. K.. (2012). Development of EST derived SSRs and SNPs as a genomic resource in Indian catfish, *Clarias batrachus*. Mol Biol Rep 39: 5921-31.

Pashley C. H., Ellis J. R., McCauley D. E. & Burke J. M. (2006) EST databases as a source for molecular markers: lessons from *Helianthus*. J Heredity 97:387-388.

Pouyaud L. & Sudarto P. E. (2009) The phylogenetic structure of habitat shift and morphological convergence in Asian Clarias, Teleostei Siluriformes: Clariidae. J Zool Sys Evol Res 47: 344-356.

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

Rozen S. & Skaletsky H.J. (2000) Primer3 on the WWW for general users and for biologist programmers In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, edited by Krawetz S & Misener S, Humana Press Totowa NJ, pp 365-386. http://primer3sourceforgenet/.

Schneider S., Roessli D. & Excoffier L. (2000) Arlequin: a software for population genetics data analysis. User manual ver 2.0. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Geneva.

Van Oosterhout C., Hutchinson W. F., Wills D. P. & Shipley P. (2004) Micro-Checker : Software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538.

Vasemagi A., Nilsson J. & Primmer C.R. (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphism for detecting signatures of divergent selection in Atlantic Salmon (*Salmo salar* L.). Mol Bio Evol 22: 1067-1076.

Vishwanath W. (2010) *Clarias magur*. In: IUCN 2012. IUCN Red List of Threatened Species. Version 2012.1 <www.iucnredlist.org>. Downloaded on 29 June, 2012 .

Volckaert F. A. M., Hellemans B. A. S. & Poyaud L. (1999) Nine polymorphic microsatellite markers in the SE Asian catfishes *Pangasius hypophthalmus* and *Clarias batrachus.* Anim Genet 30: 383-383 .

Table 1: Characteristics of 27 polymorphic EST-SSRs in *Clarias batrachus* in two natural populations.

S. No.	Locus	Accession No.	Description/ Annotation$	Primers (5' to 3')	Repeat units	Location	He	Ho	HWE (P value)*	Genetic Variation (P value)	No. of alleles, Size range (bp)
1	CbSpn1100	GR955321	Ifu.2632	F-CGGTGCACACATTCTCTCAC	(AC)8	Lucknow	0.5879	0.6957	0.5805	0.23549	5 (103-113)
				R-GTTTTAGCTGGCGCAGTTTG		Varanasi	0.4728	0.4286	0.1937		5 (105-117)
2	CbSpn1115	GR955322	Ipu.22671	F-TTCAATTTGGCTCTTTCTGG	(GTTT)9	Lucknow	0.5816	0.4091	0.0612	**0.00760**#	3 (110-118)
				R-CAAGGATTCTCATTTTCAATGTTTG		Varanasi	0.6451	0.4286	0.0242		4 (110-122)
3	CbSpn1191	GR955324	Unknown	F-TGTGCATATAGTCAACTCCAAAA	(TCTA)15	Lucknow	0.7710	0.4762	0.0047	**0.00211**#	11 (93-141)
				R-ACAGATGGGAATTGTGCTGA		Varanasi	0.8447	0.6190	0.0002*		12 (69-141)
4	CbSpn143	GR955306	Unknown	F-AAACCAGCCAAGTGTCCAGT	(AC)15	Lucknow	0.6682	0.3913	<0.0001*	0.24213	8 (87-107)
				R-GCATGGTTGTGTTCTGTGTG		Varanasi	0.7619	0.4286	<0.0001*		11 (81-109)
5	CbSpn1817	GR955333	Rho-class glutathione S-transferase [*D. rerio*]	F-CATTACTTTAACCGTATCCGTTATGA	(TC)10	Lucknow	0.6912	0.3000	<0.0001*	**0.00165**#	6 (148-158)
				R-CAGACAGGTCCGAAAGGTTG		Varanasi	0.2188	0.1429	0.0068		4 (148-154)
6	CbSpn25	GR955281	Serine/threonine-protein kinase MARK2 isoform 1	F-AATGTGTCTGGTCTGGTCTGG	(CTGAT)8	Lucknow	0.2975	0.2727	0.5399	0.11998	2 (103-108)
				R-ATCCACCTTCGGAGCACTAA		Varanasi	0.4735	0.3913	0.3210		3 (98-108)
7	CbSpn2524	GR955342	Unknown	F-CGCTGTGAGAGGAAAAGAAG	(AC)7G(CA)9	Lucknow	0.3280	0.3043	0.6238	0.30651	3 (108-124)
				R-GGGATCATATGCACACTTTACAC		Varanasi	0.2540	0.1905	0.0501		3 (108-124)
8	CbSpn2670	GR955346	Hect domain and RLD 4 [*D. rerio*]	F-CCCTGTTTCCTAATTTATTTTTGC	(TATT)5	Lucknow	0.0000	0.0000	-	**<0.0001** #	1 (144)
				R-CAGAAATCCCCCTCTTGTTG		Varanasi	0.4849	0.4783	1.0000		2 (142-146)
9	CbSpn268	GR955308	Ipu.40806	F-CTCGTGTATCTTTCATTATTTAGAAGG	(TAT)9	Lucknow	0.5875	0.2500	<0.0001*	**0.02310**#	5 (152-168)
				RGCTACATTTGTTTTCAGAAGTTAAAGC		Varanasi	0.7521	0.3182	<0.0001*		5 (153-165)
10	CbSpn839	GR955313	Hypothetical protein LOC561910 [*D. rerio*]	F-CGGAGTTAGATGTAATTATGGAGAAG	(CT)13	Lucknow	0.8730	0.5714	<0.0001*	0.17685	10 (145-167)
				R-ACCGCGTCCTCAGTCAT		Varanasi	0.8325	0.4000	<0.0001*		8 (145-159)
11	CbSpn956	GR955315	Ipu.22919	F-TCACCCAATATTAAACGATCACC	(TGA)10	Lucknow	0.7788	0.6087	0.1579	0.84608	5 (82-91)
				R-TGTGGAAAAACCCGACAGG		Varanasi	0.7656	0.6957	0.0771		5 (79-91)
12	CbSpn0449	GW397095	Unknown	F-TGCCGCTAGAATGTTCAGAT	(GT)13	Lucknow	0.5986	0.3810	0.0043	0.49351	5(150-180)
				R-ATGCATGTTTCATTTTATTGCTT		Varanasi	0.5972	0.7778	0.6762		4(154-180)
13	CbSpn0485	GR955282	myb-like, SWIRM and MPN domains 1 [*X. laevis*]	F-TCGGATGATACCGAGTTTGG	(GA)7	Lucknow	0.4163	0.1364	0.0027*	0.62986	2(104-106)
				R-CACAAAAGTACGCAGCTTCAG		Varanasi	0.4608	0.0800	<0.0001*		2(104-106)
14	CbSpn0781	GR955285	G protein-coupled receptor 126 beta 1, LOC100149780 [*D. rerio*]	F-CCAGATATTCCGCTTCTGTG	(GTT)6	Lucknow	0.5718	0.1304	<0.0001*	0.11875	3(118-124)
				R-TCGGCTGGATTTTCAGTGAG		Varanasi	0.5391	0.25	0.0035		3(118-124)
15	CbSpn0820	GW397131	translocating chain-associated membrane protein 1	F-GCCAGGAAGGACAAATCATC	(TC)6TG(TC)6	Lucknow	0.4742	0.1364	0.0010*	0.34676	2(115-117)
				R-AGGGCTGGGAATGATAAAGC		Varanasi	0.4991	0.1250	0.0001*		2(115-117)

16	CbSpn1335	GW397132	serine palmitoyl transferase 2	F-AGACCACGTACGAGGAGTGC	(CT)12	Lucknow	0.5475	0.1500	<0.0001 *	0.10805	4(124-134)
				R-GTGAAGCTTGGGAAGGAATG		Varanasi	0.6723	0.3810	0.0004*		4(124-134)
17	CbSpn1458	GW397086	hypothetical protein LOC100148940[*D. rerio*]	F-GGGAAACTAAAGAACAACCCTGAT	(AGAA)5	Lucknow	0.5406	0.0000	<0.0001*	0.85555	3(107-115)
				R-TTTCCTTTTCTTCCTTTCTCGT		Varanasi	0.5144	0.0800	<0.0001*		3(107-115)
18	CbSpn1837	GW397120	Hypothetical protein LOC553231 [*D. rerio*]	F-AAATTGCTCATCTTTCTGACG	(AT)7	Lucknow	0.0868	0.0000	0.0220	0.34678	2(115-117)
				R-CGCCATAGAGCTCTTAAATGGT		Varanasi	0.2873	0.0000	<0.0001*		2(115-117)
19	CbSpn2339	GW787427	ATPase, Na+/K+ transporting, beta 1a polypeptide [*D. rerio*]	F-GTCCAAGGTGTGCGTGTATG	(TG)8	Lucknow	0.6438	0.5500	0.2955	0.38999	4(87-93)
				R-TTACCCACACTCAGCCATGT		Varanasi	0.5946	0.4167	0.0811		4(87-93)
20	CbSpn2440	GR955341	Actin related protein 2/3 complex, subunit 4	F-TCTGATGAGAAATCCAGGTGAA	(TGA)6	Lucknow	0.6570	0.5000	0.2264	**0.01834**#	4(107-116)
				R-TTGGTTTTATCACCACCATCAG		Varanasi	0.5568	0.5455	0.6715		4(107-116)
21	CbSpn3499	GW397160	deleted in malignant brain tumors 1	F-ATTGGATTCGTTTTCATTGTCA	(TTG)11	Lucknow	0.7700	0.7000	0.0698	0.23771	5(105-117)
				R-AGACACGTGTGTTTGAGAAACC		Varanasi	0.7639	0.5833	0.0378		6(102-117)
22	CbSpn3527	GW397163	nuclear receptor coactivator partial	F-AAGAAATTAATGGTTTTGTTCCCTTA	(GT)6	Lucknow	0.5635	0.2857	0.0028*	0.8668	3(102-106)
				R-ACAGAACACGTCTCTCTCTCACA		Varanasi	0.5747	0.0870	<0.0001*		3(102-106)
23	CbSpn4566	GW840553	FcRI; High affinity immunoglobulin gamma Fc receptor	F-GCCTTGAATTAAAGCCGACA	(AATA)6	Lucknow	0.4962	0.4783	1.0000	**<0.0001** #	2(97-105)
				R-CTGTTGGTGGCAAAACGAGT		Varanasi	0.3601	0.3478	0.0966		3(97-109)
24	CbSpn4620	GW840578	cryptochrome 1a isoform 1 [*D. rerio*]	F-CTCTCACATACAAACACACACTGA	(GT)6;(TA)7	Lucknow	0.7525	0.7000	0.0032*	0.05259	7(118-142)
				R-AGACACCATGATCCCCCTTA		Varanasi	0.8300	0.3000	<0.0001*		7(118-142)
25	CbSpn4654	GW840586	cancer-up-regulated gene 2 protein	F-TGCTCATCTTTACCTTAGGACCTT	(GTGC)5	Lucknow	0.5718	0.4348	0.0008*	**0.00505**#	4(99-111)
				R-TGTTTGCAGTTACACAGTTTAGG		Varanasi	0.7786	0.5417	<0.0001*		5(99-115)
26	CbSpn Contig12-1	GW840539	hypothetical protein LOC322909 [*D. rerio*]	F-TGGCCCAGACATATCTCAAAG	(GA)7	Lucknow	0.6276	0.3478	0.0002*	**0.01817**#	3(143-149)
		GW397164		R-TGGAAGAGTGGCTTCATGTG		Varanasi	0.4688	0.3478	0.1077		3(143-149)
27	CbSpn Contig6	GR955349	Solute carrier family 12, member 6, partial	F-TGCTGCACATTTTGTCAATG	(GA)15	Lucknow	0.8209	0.6667	0.0122	**0.00575**#	7(132-152)
		GW397171		R-CATGGTCCAGGTGGAGACTT		Varanasi	0.7316	0.5652	0.0219		7(134-162)

# $p<0.05$

*Significant deviation from HWE, $p<0.0038$ after Bonferroni correction.

$ Mohindra et al. (2012)

Table 2: Cross- species amplification of twelve EST –SSR of *Clarias batrachus* in three related species.

Locus	*Clarias dussumeri** (n=2)	*Clarias gariepinus* (n=5)	*Heteropneustus fossilis** (n=5)
Cb3054	1(78)	--	--
Cb2670	--	--	--
Cb2524	2(74, 84)	--	--
Cb1817	--	--	--
Cb268	1(74)	--	1(71)
Cb839	--	--	--
Cb1115	--	--	1(111)
Cb1100	--	--	--
Cb956	2(85,91)	--	--
Cb25	1(82)	--	--
Cb1191	--	--	--
Cb143	1(136)	--	2(130-132)

* Number of alleles and allele size range (bp) in parentheses.