Indian Journal of Animal Sciences **81** (3): 302–05, March 2011

Development of 12 polymorphic EST-SSR for endangered fish, Hucho taimen (Pallas)

WANG JUN[1], KUANG YOU-YI[2], TONG GUANG-XIANG[3] and YIN JIA-SHENG[4]

Shanghai Ocean University, Shanghai 201306 China
and
Heilongjiang River Fisheries Research Institute, Harbin 100 083 China

Received: 4 October 2010; Accepted: 28 December 2010

ABSTRACT

Hucho taimen (Pallas) is an indigenous and endangered cold-water fish in China. To access the genetic diversity of wild breeding parents of this species and promote its breeding selection program, 12 polymorphic EST-SSR markers were developed from its cDNA library by $(CA)_{16}$ probe and were analyzed on 66 wild breeding parents. The number of alleles per locus (Na), expected heterozygosity (He) and polymorphism information content (PIC) ranged from 2 to 9, 0.2019 to 0.8542 and 0.1802 to 0.8312 respectively. In total, a new and economical method for EST-SSR was established. The EST-SSR markers characterized in this study will strengthen our interpretation of *Hucho taimen* and facilitate its breeding selection program to improve the endangered statue of this species.

Key words: cDNA library, EST-SSR, Genetic diversity, *Hucho taimen* (Pallas)

Hucho taimen (Pallas) is an indigenous cold-water fish with large body size and restrictedly distributed in Heilongjiang and Eerqisi River of China (Holcik *et al.* 1988). Since the problems of artificial propagation and larval fish domestication were solved, its artificial cultivation not only brings in good economic benefit, but also effectively improves the endangered statue of this species through its enhancement and releasing program. However, fry quality and quantity cannot be guaranteed due to all artificial breeding parents are derived from wild population without genetical modification. The phenomenon of slow growth, poor resistance and low survival rate are serious, which greatly affected the enhancement and releasing program. It is necessary to carry out molecular genetic studies to promote a breeding selection program. For this purpose, we developed 12 polymorphic EST-SSR markers from cDNA library by magnetic beads and used them to analysis genetic diversity of breeding parents (66 individuals), hoping that it could provide an economical and practical method for EST-SSR and lay a solid foundation for the breeding selection program.

MATERIALS AND METHODS

Genomic DNA of 66 breeding parents was extracted from fin clips (Tong 2006). Total RNA was extracted from visceral tissue with Trizol reagent by using of the CTAB method (Chang *et al.* 1993) and purified. cDNA library was constructed using RT-PCR kit according to the manufacture' instruction with some modifications, and was digested with restriction enzyme CviQI. Fragments ranged from 500bp to 900bp were retrieved using production purify kit, and then ligated to double-stranded CviQI adapter by T4 DNA ligase at 22°C for 3 h. Excess adapters were removed by washing on an Ultrafree column. The remaining digestion-ligation reaction products were used as template in a PCR amplification with linker B matching the adaptor sequence; the thermal cycler was set as follows: 94°C 3 min, 94°C 30s, 53°C 1 min, 72°C 1 min, 72°C 7 min, 17 cycles. The superfluous adaptors and primers were removed using PCR purify kit. The CviQI adapter were:

5'-GACGATGAGTCCTGAG	Linker A
TACTCAGGACTCAT-5'	Linker B

The amplification product was mixed with biotinylated probe of $(CA)_{16}$. After denaturation at 95°C for 3 min and annealing at 68°C for 1 h, streptavidin-coated beads were added to recover streptavidin-biotin complexes and were washed (2 washes by 6×SSC, 0.1jþ SDS at room temperature, 5 washes by 3×SSC, 0.1jþ SDS at 68°C and 2 washes by 6×SSC at room temperature). cDNA contained SSR was enriched after washing by 30µl of 10jþTE and was used as templates for the second round of PCR; primer and condition

Present address: [1,2,3](e mail: wangjun2008v@yahoo.com.cn) Heilongjiang River Fisherise Research Institute, Chinese Academy of Fishery Sciences, District Daoli, Harbin, China.

ICAR

Table 1. Characteristics of EST-SSRs developed for *Hucho taimen* (Pallas)

Loci	Primer sequences	Blastx top hit accession numbers	Max Identity	E-value	Closely related specie	Putative functions
HtaECA6	CAGGAGGTAGTCGCAGGTTT TGGCAGGCATTTTACTTTGG	Gb|CAF94961.1|	86%	5E–05	Tetraodon nigroviridis	unnamed protein product
HtaECA15	CCCAAACGGAAGATGAAGAC GCTAACAGGAATAGTAACAC AAACATTG	Gb|EDL87295.1|	62%	5E–06	Rattus norvegicus	isoform CRA_a
HtaECA16B	ATGTGAGCGTGTGTGTGAGC TGAGGCGTATGAGGTGAGTG	Gb|AAH95744.1|	95%	3E–26	Danio rerio	Integrin, beta-like 1
HtaECA29	AATGTGCTTGAGTTGATGCCTG TATATCCGTCGACCTTGAATCTG	Gb|EDL09599.1|	51%	7E–08	Mus musculus	mCG147311
HtaECA38	CTGTGTCGTGGGCTACTTTT ATTCATCATTCCCCTGGTCA	Gb|AAI54019.1|	38%	6E–04	Danio rerio	Zgc:171859 protein
HtaECA47	TCGGAATAAGCCCAAAACTG TCCCCCGTGACTAAAAACAT	Gb|AAF16687.1|	51%	0.007	Homo sapiens	AF111848_1 PRO0529
HtaECA67	AGGGGATGAGGGTCAAAAAG CCAGTCGCAAAACACAGAAG	Gb|ACI33868.1|	87%	2E–11	Salmo salar	Proliferation-associated protein
HtaECA79	TCTGTCATTTCACCGTTCCA CCACCCGACTACCACTTCAG	ref|XP_001993678.1|	40%	0.27	Drosophila	GH21031
HtaECA82	TCTGTTTGCTGGCTTATTATGTT GGGTTGGATCAGTAGAGGTGT	Gb|CAD97815.1|	96%	3E–10	Homo sapiens	hypothetical protein
HtaECA90	AGGGGATGAGGGTCAAAAAG CAACCGAGGAGAGCAACAAG	Gb|ACI33868.1|	96%	3E–07	Salmo salar	Proliferation-associated protein
HtaECA91	AGGGGATGAGGGTCAAAAAG CAACCGAGGAGAGCAACAAG	Gb|ACI33868.1|	96%	1E–06	Salmo salar	Proliferation-associated protein
HtaECA100	AGGGGATGAGGGTCAAAAAG CCAGTCGCAAAACACAGAAG	Gb|ACI33868.1|	97%	6E–31	Salmo salar	Proliferation-associated protein
HtaECA104	TGACCCCTTAGCCTTCATTTAC GCCCCTCCAAGTGTCCAGTA	ref|NP_001019243.1|	75%	2E–04	Danio rerio	integrin, beta-like 1
HtaECA106	CCCCACTTGCTGCTTCTCTA TCTTCCTCCCACACACACAC	ref|NP_001070198.1|	88%	2E–09	Danio rerio	hypothetical protein
HtaECA114	CAAGGCACAAAGTCACCAAA TCATCCATCCTCACCATCTG	ref|NP_001134186.1|	89%	1E–38	Salmo salar	Serine/threonine-protein kinase
HtaECA121	TCAGCACAAACTACAATCATCACTA TCACAGCCCCCTTCATTATC	ref|NP_001070198.1|	90%	5E–06	Danio rerio	hypothetical protein
HtaECA125	GTGGGAGTTGGGACAAAAGA CCCATTCATCGCCATTTTAC	None	None	None	None	None
HtaECA131	AGGGCTTGGTTTCCATTAGT TGTCCTCAACGCTTCTCTATG	dbj|BAC33074.1|	52%	0.009	Mus musculus	unnamed protein product
HtaECA22	ATTTTCAGCCCATAACATCAAC GCAATAAATGAGAAGGAAATGG	emb|CAG05028.1|	52%	0.013	Tetraodon nigroviridis	unnamed protein product
HtaECA50	ATGTTTTGGGAGAATACCTGAG ATAGGCAGAGGCAGCACAG	gb|AAH37320.1|	37%	3.9	Homo sapiens	FBXW7 protein
HtaECA79	TCTGTCATTTCACCGTTCCA CCACCCGACTACCACTTCAG	ref|XP_001993678.1|	40%	0.27	Drosophila	GH21031
HtaECA27	CCTTAGCGAGTTGATGATGTGG GGCGGGAGTTCCTTCTTAGC	emb|CAF93229.1|	61%	5E-05	Tetraodon nigroviridis	unnamed protein product
HtaECA16	TGTGTGTGTGTGTGTGTGTGA GAGAGTGGGTGGGTGTGTGT	Gb|AAH95744.1|	95%	3E-26	Danio rerio	Integrin, beta-like 1
HtaECA16C	CACAGAGAGCGACAGAGACAG TGAGGTGAGTGGAGTGAGG	Gb|AAH95744.1|	95%	3E-26	Danio rerio	Integrin, beta-like 1

were the same as stated above but 30 cycles instead of 17 cycles. The PCR product was ligated to pMD18-T vector (TaKaRa) at 16°C for 3 h, and then transformed into competent *E.coli* $DH5_{\alpha}$ that activated by $CaCl_2$ (Sambrook and Russell 2001). After cultured on solid Luria-Bertani plates containing 100μg/mL ampicillin, 1100 recombinants were picked out for the secondary detection by PCR amplification with $(CA)_{10}$ primer and M13" primer. Finally,

ICAR

Table 2. Genetic parameters of wild populations of *Hucho taimen* (Pallas) analyzed by EST-SSRs

Loci	PCR product size range(bp)	Na	Ho	He	Fis	PIC	P equilibrium
HtaECA67	280–306	3	0.7167	0.5368	–0.3462	0.4344	0.0290**
HtaECA6	194–235	7	1.000	0.7836	–0.2859	0.7508	0.0000**
HtaECA82	244–260	2	0.4833	0.5041	0.0331	0.3749	0.7480
HtaECA16B	174–256	9	0.9355	0.8542	–0.1041	0.8312	0.0000**
HtaECA15	187–216	3	0.4603	0.5013	0.0745	0.4294	0.7042
HtaECA121	220–284	8	0.9048	0.8128	–0.1220	0.7809	0.0034**
HtaECA106	187–215	4	0.9524	0.6847	–0.4021	0.6207	0.0000**
HtaECA29	306–341	4	0.9683	0.7328	–0.3318	0.6777	0.0000**
HtaECA131	171–199	4	0.9848	0.6827	–0.4535	0.6290	0.0000**
HtaECA47	180–184	2	0.2258	0.2019	–0.1273	0.1802	0.3359
HtaECA91	135–162	4	0.9697	0.5759	–0.6967	0.4823	0.0000**
HtaECA90	136–162	4	1.000	0.6458	–0.5602	0.5691	0.0000**
Mean		4.5	0.8001	0.6264	–0.2768	0.5634	

Note: Na, numbers of alleles per locus detected, Ho, observed heterozygosity; He, expected heterozygosity; Fis, fixation index. P, Hardy-Weinberg equilibrium index; *significant deviation from equilibrium; **extremely significant deviation from equilibrium.

417(37.9%) positive clones were acquired and 200 of them were sent for sequencing. Sequences of linkers and vectors were removed by VLDeleter software.

Repeat motifs of sequencing results were found out by TRF software with dinucleotides ≥ 5. Primers were designed by using of Primer3 software with default setting (Steve Rozen *et al.* 2000). Putative functions of the sequences were identified by blasting them to the non-redundant (nr) database of NCBI Database (Altschul SF 1990). Annealing temperature and polymorphism of designed primers were examined using wild population of 66 breeding parents. The PCR amplification was carried out in total volumes of 15µl containing: 100ng genomic DNA, 10uM of each primer, 0.5 mM dNTP, 10× reaction buffer and 1U taq polymerase. The thermal cycler was set as follows: 94°C 3 min, 94°C 30s, annealing temperature 30s, 72°C 30s, 72°C 7 min, 26–28 cycles. Allelic size was estimated using GeneRuler 100bp DNA ladder by GEL-PRO 4.5 software. The numbers of alleles per locus (Na), observed heterozygosity (Ho), expected heterozygosity (He) and H-W equilibrium test was analyzed by PopGen3.2 software (Tech Sigh 2000). Polymorphism information content (PIC) (Bostein *et al.* 1980) was calculated with PIC_Calc 0.6 software by following formula:

$$PIC = 1 - \sum_{i-1}^{n} P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2P_i^2 P_j^2$$

where, P_i, P_j, frequencies of alleles, n, the loci number.

RESULTS AND DISCUSSION

In total, 143 (71.5%) of sequencing results were detected containing repeat motifs. However, only 24(16.8%) primers were successfully designed due to the inadequate sequences flanking on either ends of the SSRs or inappropriate base composition. All of the 24 sequences have showed high similarity against database of EST others while 23 sequences against database of nr protein. The characteristics of microsatellites and primers sequence are shown in Table 1. 22 primers amplified suitable PCR products, 12 of them showed clear polymorphism, with Na, Ho, He and PIC ranging from 2 to 9, 0.2258 to1, 0.2019 to 0.8542 and 0.1802 to 0.8312 respectively (Table 2). Mean PIC was 0.5634 suggesting that the genetic diversity of the wild breeding parents population was relatively high and selective breeding on this population has broad prospect. Significant linkage disequilibrium among all pairs of loci was not found ($P<0.05$). However, All locus were significant deviations from H-W equilibrium ($P<0.05$) except HtaECA82, HtaECA15 and HtaECA47. In this study, a new method to develop EST-SSR was established especially for the species without in-depth study or rich source database of ESTs. Two hundred EST sequences and 12 polymorphic EST-SSR markers characterized in present study will strengthen our interpretation of genome of *Hucho taimen* and facilitate its breeding selection program.

ACKNOWLEDGMENTS

This work was financially supported by The Central-Level Non-profit Scientific Research Institutes Special Funds (2008HSYZX-SJ-08) and Non-profit Scientific Research Funds (201003055).

REFERENCES

Altschul S F, Gish W, Miller W.1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–10.

Bostein D, White R L, Sckolnick M. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 317–31.

ICAR

Chang S, Puryear J and Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporters* **11**: 113–16.

Holcik J, Hensel K, Nieslanik J.1988. *The Eurasian Huchen, Hucho* Hucho: *Largest Salmon of the World*. Kluwer Academic Publishers, Hingham, assachusetts. Retrieved from: http://onlinelibrary.wiley.com/doi/10.1002/iroh.19890740617/.

Kuang You-Yi, Sun Xiao-Wen, Yin Jia-Sheng. 2009. Batch implement for contamination sequences in DNA sequences. China, ZL 200710144577.X, 2009-11–18.

Sambrook and Russell. 2001. Molecular Cloning: *A Laboratory Manual*. 3rd edn. Cold Spring Harbor Laboratory Press.

Steve Rozen, Helen J and Skaletsky. 2000. Primer3 on the www for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. pp 365–386. (Eds) Krawetz S, Misener S. Humana Press, Totowa, New Jersey.

Tong Guangxiang, Kuang Youyi and Yin Jiasheng. 2006. Isolation of microsatellite DNA and analysis on genetic diversity of endangered fish, Hucho taimen(Pallas). *Molecular Ecology Notes* **6**: 1099–1101.

Toth G, Gasparl Z and Jurka J. 2000. Microsatelites in different eukaryotic genome, survey and analysis. *Genome Research* **10**: 1967–81.

ICAR