

**Abstract**—Twenty homologous microsatellite markers, or simple sequence repeats (SSRs), were developed for blackfin tuna (*Thunnus atlanticus*) through the use of a direct *Seq-to-SSR* approach. The number of alleles per locus ranged between 5 and 26, and the expected heterozygosity ranged between 0.640 and 0.969. Three loci displayed significant departure from Hardy-Weinberg equilibrium expectations, likely reflecting occurrence of null alleles. Another locus showed consecutive alleles that differed by one base pair only. Consequently, this locus may be prone to elevated rates of scoring errors. The remaining 16 loci will be valuable for studies in conservation genetics of blackfin tuna.

## Development and characterization of microsatellite markers for blackfin tuna (*Thunnus atlanticus*) with the use of Illumina paired-end sequencing

Luca Antoni<sup>1</sup>

Patricia L. Luque<sup>1</sup>

Kaylie Naghshpour<sup>1</sup>

Lionel Reynal<sup>2</sup>

Eric A. Saillant (contact author)<sup>1</sup>

Email address for contact author: eric.saillant@usm.edu

<sup>1</sup> Gulf Coast Research Laboratory  
The University of Southern Mississippi  
703 East Beach Drive  
Ocean Springs, Mississippi 39564

<sup>2</sup> Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)  
Unité Biodiversité et Environnement de la Martinique  
79 Route de Pointe-Fort  
97231 Le Robert  
Martinique, France

The blackfin tuna (*Thunnus atlanticus*) is a small tuna widely distributed in tropical and subtropical waters of the western Atlantic Ocean from the mid-Atlantic region of the East Coast of the United States south to northern Brazil (Collette and Nauen, 1983). Although this species is exploited by fisheries in several countries, management of this resource in the United States and abroad is almost inexistent and, in particular, it is not known to date whether multiple stocks of this species occur within its recorded distribution range (Mathieu et al., 2013).

Tagging studies conducted in the island nations of St. Vincent and the Grenadines and Bermuda have indicated that blackfin tuna exhibit some degree of site fidelity for a significant proportion of tagged fish that have been recaptured in close proximity of their tagging location, sometimes over multiple years (Luckhurst et al., 2001; Singh-Renton and Renton, 2007). However, long distance movements also were reported for some of the individuals tagged in the study of Singh-Renton and Renton (2007). In

such a situation, the analysis of genetic variation at molecular markers is expected to provide valuable information on the structure of stocks to develop sustainable management of fisheries (Carvalho and Hauser, 1994). Microsatellites are the most widely used markers for genetic studies of exploited fishes; however to date homologous markers for the blackfin tuna are not available.

The objective of this work was to develop homologous microsatellite markers, or simple sequence repeats (SSRs), for studies in conservation genetics of this species. The application of next-generation sequencing technologies greatly enhances microsatellite discovery through direct screening of short sequences of genomic DNA, hereafter *reads*, for repeat arrays without the need for cost-prohibitive steps, such as cloning of genomic libraries and screening of individual clones through Sanger sequencing. The *Seq-to-SSR* method of Castoe et al. (2012) is based on direct screening of unassembled sequencing reads, an approach that further increases the cost effectiveness of the

Manuscript submitted 7 January 2014.  
Manuscript accepted 25 August 2014.  
Fish. Bull. 112:322–325 (2014).  
doi:10.7755/FB.112.4.8

The views and opinions expressed or implied in this article are those of the author (or authors) and do not necessarily reflect the position of the National Marine Fisheries Service, NOAA.

procedure. In our study, this method was applied to rapidly identify microsatellite loci in the blackfin tuna.

## Materials and methods

Genomic DNA from fin tissue of one blackfin tuna specimen collected offshore of the Louisiana coastline was extracted by using a standard phenol-chloroform protocol (Sambrook et al., 1989). An Illumina<sup>1</sup> paired-end library, which allows sequencing of both ends of DNA fragments, was prepared and sequenced on an Illumina HiSeq 2000 platform (Illumina, San Diego) according to methods described by Castoe et al. (2012). Reads were quality controlled and trimmed for low-quality data (phred scores <30). In the program PAL\_FINDER, vers. 0.02.03, a Perl script developed by Castoe et al. [2012] and available at <http://sourceforge.net/projects/palfinder/>, 5,874,294 reads were screened for microsatellite arrays that contained a minimum of 12 repeats. Use of a minimum of 12 repeats, in our experience, ensures that the selected microsatellite loci are likely to be polymorphic. The search in PAL\_FINDER was continued until 45 dinucleotide loci and 5 tetranucleotide loci were discovered, requiring screening of 286,240 reads and the whole database (5,874,294 reads) for dinucleotides and tetranucleotides, respectively.

Specific polymerase chain reaction (PCR) primers were designed with the open-source software Primer3 (Untergasser et al., 2012; Koressaar and Remm, 2007; source code available at <http://sourceforge.net/projects/primer3/>). The designed primers were tested for amplification success and scoring reliability and then evaluated at different annealing temperatures using blackfin tuna samples provided by the Louisiana Department of Wildlife and Fisheries. Samples had been collected offshore of the Louisiana coastline during the spring of 2013. Amplification success of the candidate loci during PCR was tested by assaying 6 specimens. PCR reactions were performed in a total volume of 5.6  $\mu$ L that contained 7–13 ng of genomic DNA, 2.2 pmol of each primer, 1.1 nmol of dNTPs (Promega Corp., Madison, WI), 8.4 nmol of MgCl<sub>2</sub> (Promega), 0.28 U of Taq polymerase (Promega), and 1X of buffer (Promega). Amplification by PCR consisted of an initial denaturation at 95°C for 5 min, 35 cycles of 95°C for 30 s, annealing temperature ( $T_a$ ) for 30 s, 72°C for 45 s, and a final extension of 15 min at 72°C.  $T_a$  was 62°C for the first 7 cycles, 60°C for the following 7 cycles, and 56°C for the remaining 21 cycles. The obtained PCR products were evaluated through electrophoresis on high-resolution NuSieve GTG Agarose gels (Lonza Group, Basel, Switzerland).

Loci that showed consistent amplification success and polymorphic PCR products were selected for fur-

ther testing, and the 5' end of 1 of the 2 primers was labeled by using one of the fluorescent dyes 6-Fam, Hex, or Ned to allow detection and scoring on an automated DNA sequencer. Evaluation of scoring reliability of the tested loci was based on DNA samples from 8 blackfin tuna specimens. The optimal  $T_a$  for amplification of each locus was then determined during PCR reactions by using DNA from 4 blackfin tunas, and was carried out at 6 annealing temperatures held constant through the 35 amplification cycles and ranging from 52°C to 62°C. PCR products were run on an ABI Prism 377 96-lane DNA Sequencer (Applied Biosystems, Life Technologies, Carlsbad, CA). The obtained electropherograms were analyzed with Applied Biosystems GeneScan software, vers. 3.1.2 (Life Technologies), and alleles were scored in Applied Biosystems Genotyper software, vers. 2.5 (Life Technologies). The polymorphic loci that could be scored reliably were characterized statistically on the basis of the genotypes of 35 blackfin tuna specimens provided by C. Pau and L. Reynal, both of IFREMER, La Martinique. Specimens had been caught offshore of the island of Martinique in the summer of 2013.

## Results

Amplification tests were conducted on 50 primer pairs, 45 potentially amplifying dinucleotide microsatellites, and 5 amplifying tetranucleotide microsatellites. Twenty-four loci were amplified consistently across the tested specimens and were all labeled with fluorescent dyes for further evaluation and optimization of annealing temperature. Twenty loci (19 dinucleotides and 1 tetranucleotide) gave scorable PCR products and are described in Table 1. The number of alleles ( $A$ ), expected heterozygosity ( $H_e$ ), and inbreeding coefficient ( $F_{IS}$ ) were calculated with the software FSTAT, vers. 2.9.3.2 (Goudet, 1995). Per locus estimates ranged from 5 to 26 for  $A$ , from 0.640 to 0.969 for  $H_e$ , and from -0.003 to 0.268 for  $F_{IS}$ . Conformance of genotypic proportions to Hardy-Weinberg (H-W) equilibrium expectations was tested with exact tests in the software GENEPOP, vers. 4.1 (Raymond and Rousset, 1995; Rousset, 2008). Genotypic proportions did not depart significantly from H-W equilibrium expectations, except for the loci BT6, BT27, BT47, and BT91. The departure at locus BT27 was not significant after Bonferroni correction (Rice, 1989). Analyses in Micro-Checker, vers. 2.2.3 (Van Oosterhout et al., 2004) revealed possible occurrence of null alleles at loci BT6, BT47, and BT91. There was no evidence of scoring artifacts at locus BC27.

## Discussion

On the basis of the significant departure of genotypic proportions from Hardy-Weinberg equilibrium expectations and the inference of null alleles at loci BT6,

<sup>1</sup> Mention of trade names or commercial companies is for identification purposes only and does not imply endorsement by the National Marine Fisheries Service, NOAA.

Table 1

Summary data for 20 microsatellites developed for the blackfin tuna (*Thunnus atlanticus*).  $T_a$ : annealing temperature; A: number of alleles;  $H_e$ : expected heterozygosity;  $F_{IS}$ : inbreeding coefficient;  $P_{H-W}$ : probability of conformance to Hardy-Weinberg (H-W) equilibrium expectations. Values in bold represent significant departures from H-W expectations after Bonferroni correction. An asterisk (\*) indicates that the sequence of the microsatellite array could not be determined in its entirety because of insufficient overlap of the 2 paired-end reads (the reported number is the highest number of repeats contained within 1 of the 2 paired-end readings).

Locus	Repeat motif	Primer sequence (5'-3') label	$T_a$ (°C)	A	Allele size range (base pairs)	$H_e$	$F_{IS}$	$P_{H-W}$
BT4	(CA) <sub>2</sub> TA(CA) <sub>13</sub>	F: GGTTGAAAGTGCTTCAGGTGTAA (FAM) R: CCTGCCACAGAAATGTCTGAA	58	17	134-176	0.905	0.038	0.0972
BT6	(CA) <sub>17</sub> *	F: ACAGCAGACTGTCAAAACACAAA (HEX) R: CTTTGTGGCAGAAAGCTTTGA	62	30	142-242	0.969	0.145	<b>0.0006</b>
BT11	(CA) <sub>13</sub> *	F: AACCTTAACCACTGACCCAAA (FAM) R: CTGGTTAGAGCTTTGTCAATTG	62	22	198-258	0.949	0.039	0.8843
BT18	(CA) <sub>12</sub>	F: CCATAAAAACACTGCCTGTCAGA (HEX) R: AGAAACGCTTTGGATGCAACT	54	15	96-136	0.923	-0.017	0.2059
BT20	(CA) <sub>14</sub> TA(CA) <sub>4</sub> *	F: GCGAGACTTGTGCCCTGACTT (HEX) R: ATGGCATGGTAGGTGATGTT	62	20	172-224	0.929	0.05	0.2354
BT22	(CA) <sub>18</sub> *	F: TATCTGATAAAGCAGGGAAGACC (HEX) R: AAATTTGCAAGTGGCTCTAA	56	18	163-243	0.903	0.055	0.1326
BT27	(CA) <sub>17</sub>	F: CCACATTAACAATCCGTGATCC R: GCGTTTAGGTAAGTGTGGTGT (FAM)	62	14	233-273	0.89	0.052	0.0197
BT29	(CA) <sub>14</sub> *	F: CACGACAGGCCATAATCAATC R: TTGAGGCCCTACAGATGTCAGA (NED)	62	19	162-208	0.915	0.001	0.5591
BT31	(CA) <sub>25</sub> *	F: CATGACCCAGGAAACACCTT R: AAATGCCCTTGATGCTGTATG (FAM)	58	12	365-403	0.89	0.042	0.2107
BT37	CACC(CA) <sub>12</sub>	F: TGATCAGTATGAACGGTCCCTT (HEX) R: GTGCTTGTCTCCTGAAAGCTC	62	5	304-316	0.646	0.16	0.0503
BT47	(GA) <sub>15</sub>	F: GGATGGTGCACAGGTTCTC (FAM) R: CAGGAACATCAACAGGAATGA	54	13	126-156	0.776	0.263	<b>0.0008</b>
BT67	(CA) <sub>12</sub>	F: CAGGCTGAGGATGGTGTAC R: TGTATCACACCGCCAGGTAA (FAM)	62	8	137-163	0.67	-0.086	0.108
BT68	(CA) <sub>13</sub>	F: GTAAGCACCAAGCCTGTGAA (HEX) R: AGCAGGCTGAGCTTAACTGG	62	10	95-113	0.815	-0.003	0.75
BT71	(CA) <sub>3</sub> CG(CA) <sub>20</sub>	F: TTAAACGAGCTTACTTGTCTCA R: GTGAGGTCACGCTTTTATGC (HEX)	62	26	225-305	0.943	0.061	0.0698
BT73	(CA) <sub>26</sub>	F: GCCTGTACTCCGAGGTAA (FAM) R: TGTGTCTGTAAACCCAGCTT	62	25	229-387	0.952	0.011	0.5812
BT81	(CA) <sub>13</sub> TACA	F: AGGCTCAAGGTCACATC (FAM) R: GGGCCAAAGTGAACAGTAAA	62	12	210-232	0.852	0.137	0.3411
BT83	(CA) <sub>13</sub>	F: AGGCTGACTACAAATCTGTT (HEX) R: CAGTGGTGAATAACAGCATGTC	62	7	144-158	0.64	-0.16	0.898
BT88	(CA) <sub>19</sub> *	F: AGCTTGGTGAATAATGCAAGAC (FAM) R: AGGAGCAGATGCTCTCAGGA	58	19	151-197	0.924	-0.051	0.6784
BT91	(GATA) <sub>12</sub> *	F: CAGACCTTAGTGTCTGCT (NED) R: CGCTTCCAGGTTAGTTGTT	62	20	155-211	0.937	0.268	<b>0.0005</b>
BT95	(CA) <sub>12</sub> *	F: ATTCTGCAAAAGGAGGTGTT R: TCTTTCTGCTGACCCGATGTG (HEX)	58	14	190-234	0.897	0.013	0.7854

BT47, and BT91, we caution against the use of these 3 microsatellites for population genetic studies of the blackfin tuna. Genotypic frequencies at locus BT27 did not depart significantly from Hardy-Weinberg equilibrium expectations, but analyses in Micro-Checker indicated that null alleles may occur at this microsatellite. This marker, therefore, will need to be evaluated further with larger sample sizes to determine its suitability for studies of population genetics of the blackfin tuna. Finally, locus BT71 showed occurrence of rare shifts of one base pair (i.e., consecutive alleles differed by only one base pair, departing from the expected pattern of variation at microsatellites). Because of the limited resolution of acrylamide gels, the ability to reliably score alleles that differ by one base pair is challenging and, consequently, this locus may be prone to elevated rates of scoring errors. The microsatellite markers developed during our study will be available for conducting studies of the genetic stock structure of blackfin tuna, investigations that are needed to assist in the design of management plans for the sustainability of this species.

### Acknowledgments

We thank K. Jones for help with the implementation of PAL\_FINDER software and C. Pau for assistance with collection of blackfin tuna specimens. This work was supported by the Moored Fish Aggregating Device in the Lesser Antilles (MAGDELESA) project cofounded by IFREMER and the Fonds européen de développement régional (FEDER) within the framework of the INTERREG Caraïbes Programme (contract no. 13/1210870). K. Naghshpour was supported by the Tidelands Trust Fund Program of the Mississippi Department of Marine Resources (award no. S-13-RIP-USM/GCRL-01). Views expressed in this article are those of the authors and do not necessarily reflect views of the sponsors.

### Literature cited

- Carvalho, G. R., and L. Hauser.  
1994. Molecular genetics and the stock concept in fisheries. *Rev. Fish Biol. Fish.* 4:326–350.
- Castoe, T. A., A. W. Poole, A. P. J. de Koning, K. L. Jones, D. F. Tomback, S. J. Oyler-McCance, J. A. Fike, S. L. Lance, J. W. Streicher, E. N. Smith, and D. D. Pollock.  
2012. Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE* 7(2):e30953. doi:10.1371/journal.pone.0030953
- Collette, B. B., and C. E. Nauen.  
1983. FAO species catalogue, vol. 2. Scombrids of the world. An annotated and illustrated catalogue of tunas, mackerels, bonitos and related species known to date. FAO Fish. Synop. 125, 137 p. FAO, Rome.
- Goudet, J.  
1995. FSTAT (vers. 1.2): a computer program to calculate F-statistics. *J. Hered.* 86:485–486.
- Luckhurst B. E., T. Trott, and S. Manuel.  
2001. Landings, seasonality, catch per unit effort and tag-recapture results of yellowfin tuna and blackfin tuna at Bermuda. *Am. Fish. Soc. Symp.* 25:225–234.
- Koressaar T., and M. Remm.  
2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23:1289–1291.
- Mathieu, H., C. Pau, L. Reynal, and D. Theophille.  
2013. What do we know about blackfin tuna (*Thunnus atlanticus*)? In Proceedings of the 65th Gulf and Caribbean Fisheries Institute; Santa Marta, Colombia, 5–9 November 2012, p. 245–249. Gulf and Caribbean Fisheries Institute, Fort Pierce, FL.
- Raymond, M., and F. Rousset.  
1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86:248–249.
- Rice, W. R.  
1989. Analyzing tables of statistical tests. *Evolution* 43:223–225.
- Rousset, F.  
2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* 8:103–106.
- Sambrook, J., E. F. Fritsch, and T. Maniatis.  
1989. *Molecular cloning: a laboratory manual*, 2<sup>nd</sup> ed., 1659 p. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Singh-Renton, S., and J. Renton.  
2007. CFRAMP's large pelagic fish tagging program. *Gulf Caribb. Res.* 19(2):99–102.
- Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen.  
2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40(15):e115. doi:10.1093/nar/gks596
- Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley.  
2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4:535–538.