Development and characterisation of tri- and tetra-nucleotide polymorphic microsatellite markers for skipjack tuna (*Katsuwonus pelamis*)

SudathT. Dammannagoda, Vincent Chand, and Peter B. Mather
Biologicalsciences, Faculty of Science and Engineering, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia

Corresponding author: Sudath T. Dammannagoda s.dammannagoda@qut.edu.au, Tel: +61 7 31384152, Fax: +61 7 31381535

Abstract

Skipjack tuna (*katsuwonus pelamis)* (SJT) is the largest tuna fishery in all the major oceans around the world, and the largest marine fishery in Sri Lanka. Knowledge of genetic population structure and effective population size of SJT in the Indian Ocean and other major oceans, however, is still lacking for better management practices and conservation strategies. We developed microsatellite genetic markers using SJT around Sri Lanka in the Indian Ocean, and characterise one tri- and seven tetra-nucleotide microsatellite loci isolated from enriched genomic libraries from SJT, to provide tools for addressing both conservation and fisheries management questions. An analysis of these eight microsatellite markers in two populations of SJT from eastern Sri Lanka (n = 44) and the Maldives Islands (n = 53) showed that all eight microsatellites were polymorphic with an average number of alleles per locus of 11.80 (range 5-27). Expected heterozygosities at marker loci ranged from 0.450 to 0.961. These markers are being used currently to characterise population structure and extent of natural gene flow in SJT populations from the eastern and western Indian Ocean. No significant linkage disequilibrium was detected among any loci pairs.

Keywords: enrichment, genetic diversity, microsatellites, population structure, skipjack tuna**.**

Introduction

Skipjack tuna (SJT), *katsuwonus pelamis* (Linnaeus, 1758), are widely distributed across all major oceans around the world and constitute a major food commodity both regionally and internationally. In addition, while it has not been widely recognised, SJT constitutes the largest tuna fishery worldwide. In 2009, global SJT catch had reached 2,599,681 metric tonnes (mt), and comprised 42% of the total global tuna catch (FAO 2011). Based on life history characteristics and fishery data SJT are currently considered, like most tuna species, to constitute a single reproductive unit and hence are managed essentially as a single stock in all oceans around the world by the respective international commissions (i.e. IOTC-Indian Ocean Tuna Commission; WCPFC-Western and Central Pacific Fisheries Commission; IATTC- Inter-American Tropical Tuna Commission; and ICCAT-International Commission for the Conservation of Atlantic Tuna). Some recent studies of SJT that used both mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) microsatellite markers however, detected two divergent mtDNA clades and fine-scale geographical population structure in the northwestern Indian Ocean (Dammannagoda et al. 2011). Information on wild population structure will be critical for conservation management efforts for SJT, but highly sensitive genetic markers are essential for determining wild stock structure as SJT have a high capacity to disperse long distances and wild populations are still large in some regions.

Although SJT comprise the largest tuna fishery in the world, only a few genetic studies have been conducted to assess their wild stock structure. Studies have mainly concentrated in the Pacific and Atlantic Oceans. A population genetic study of SJT carried out at an oceanic scale in the Atlantic and Pacific Oceans that employed mtDNA RFLP (Restriction Fragment Length Polymorphism) data could not detect any differentiation between Atlantic and Pacific Ocean populations (Graves et al. 1984), or from the Atlantic and Indian Oceans (Ely et al. 2005). Another study of SJT population structure could differentiate samples from India and Japan using mtDNA RFLP markers (Menezes

et al. 2005), it failed however, to differentiate the Japanese sample and a SJT sample from the western coast of India using microsatellite markers (Menezes et al. 2008).

Microsatellites commonly provide highpervariable markers that can provide the sensitivity necessary to detect even weak population structure where it exists. Although several studies have developed and characterised microsatellite markers for tuna species (e.g. Atlantic bluefin tuna; Broughton and Gold, 1997; Takagi et al. 1999; McDowell et al. 2002; Clark et al. 2004), there has been only a single report of microsatellite primers developed for SJT (i.e. Menezes et al. 2008), and they only reported on development of dinucleotide markers. Dinucleotide microsatellite markers frequently suffer from stutter bands that are more prone to genotype scoring errors than equivalent tri- and tetranucleotide repeat loci (e.g. McDowell *et al*. 2002). Scoring problems associated with dinucleotide markers can be particularly problematic when relatively large numbers of alleles are present at marker loci, in addition to the large sample sizes that are often required to adequately estimate real allele frequencies. In the current study we focussed our attention on developing tri- and tetra-nucleotide microsatellite markers for SJT because in general, they should result in fewer problems with stutter bands and they commonly have fewer number of alleles per locus than comparable dinucleotide repeat microsatellites. Tri- and tetra-nucleotide microsatellites, however, are much less abundant in the nDNA genome compared to dinucleotide microsatellites, making development of tri- and tetra-nucleotide microsatellites a difficult task.

Methodology

Genomic library development

A single tri- and seven tetra-nucleotide microsatellite markers were developed for SJT after generating a microsatellite library using a radio isotopic method (Chand et al. 2005) at the molecular genetics labs, Queensland University of Technology, Brisbane, Australia. Genomic DNA

from SJT samples that were collected around Sri Lanka was extracted using a salt extraction method (Miller et al. 1998), and 10µg of DNA was digested with DpnII (5' GATC 3') and Sau3A I (3' CTAG 5'). Digested, purified DNA (300-700 bp) was ligated to a plasmid vector pUC 18 Bam I/BAP (Amersham Pharmacia Biotech). Ligated DNA with the vector was transferred to *E.coli* competent cells using heat shock. Competent cells with ligated DNA were then cultured on LB/AMP/X-Gal/IPTG plates, and positive clones were identified. Colonies with positive clones were transferred to Hybond^{+} nylon membranes and DNA was fixed and hybridized with a mixture of radioactive labeling oligo-nucleotide probes $(CAC)_8$, $(CAG)_8$, $(GACA)_6$, $(GATA)_6$.

Clones with microsatellite inserts were identified by exposing hybridized membranes on X-ray film. Positive clones were selected from the plates, and grown in Terrific Broth/Ampicillin 50mg/ml (TB/Amp) solution. DNA was then extracted from grown out positive clones using a miniprep protocol. RNAase treated positive clones were sequenced using M13F primer (5' GTA AAA CGA CGG CCA GT '3) at the "Australian Genome Research Facility" (AGRF) (http//www.agrf.org.au) and checked for microsatellite repeats discarding short repeats and duplicated clones. Primers were designed using the PRIMER3 programme (Rozen and Skaetsky, 2000).

Primer designing and optimisation, and screening of populations

Approximately 200 positive SJT clones were identified of which 80 were sequenced. Based on sequence analysis, 17 unique primer pairs were designed and evaluated for polymorphism. Clone sequences were deposited in GenBank with accession numbers from HM631812 to HM631828, respectively. Of the 17 primer pairs screened, 10 primer pairs produced reliable microsatellite products. Of these 10 loci, two could not be scored unambiguously due to amplification of spurious bands. The remaining eight loci were polymorphic and considered to be suitable for screening variation in SJT populations (Table 1). Suitability of these eight microsatellite markers to detect

genetic variation of SJT populations was tested by screening 44 individuals from a single site in the east of Sri Lanka (Kalmunei - KM), and also 53 individuals from the Maldive Islands (Male - MD) as a remote group, at the eight loci. The PCR reaction mix consisted of ~50ng/µl DNA 1µl, 1.25µl of 10X PCR buffer (Roche), 0.25µl of 25mM $MgCl_2$, 0.5µl of 10mM dNTP (Roche), 0.5µl of each 10mM forward and reverse primers, 0.1µl of Taq (Roche) and ddH_2O to a final volume of 10 µl. Forward primers were labelled with Hexa fluorescent dye (GeneWorks, Hindmarsh, SA, Australia) for visualisation. Amplification consisted of an initial denaturation of 4 minutes at 95^0C, then 29 cycles of (i) 95^0C for 30 seconds, (ii) at appropriate annealing temperature (Table 1) for 30 seconds, (iii) 72 ^{0}C for 30 seconds; and a final extension step at 72^0C for 8 minutes. PCR products were electrophoresed in 0.6X TBE 6% acrylamide gels (100µm thick) on a GS2000 Genetic analyser (Corbett Research) According to the Corbett manual, with a 50-350 bp sizing standard (Tamra-T350) to determine allele size. Resulting products were scored using OneDscan software (Scanalytics Inc.). Summary data for the 17 microsatellite loci developed are presented in Table 1.

Statistical analysis

Raw microsatellite data were summarised into allele frequencies for each locus at two sites using the software CONVERT (Glaubitz, 2004). This software was also used to transform data into other data formats for subsequent analyses using other statistical software packages. Microsatellite data were checked for presence of null alleles, large allele dropout and errors in scoring using Micro-checker software version 2.2.3 (Oosterhout et al. 2004). Conformance of genotypic frequencies at each locus to Hardy-Weignberg equilibrium (HWE) expectations was estimated with Arlequin version 3.5.1.3 software (Schneider et al. 2005) with significance of deviations in observed vs expected heterozygosity tested using Exact tests (Guo and Thompson 1992). The possibility of linkage disequilibrium (LD) among loci was investigated using the method of Slatkin and Excoffier (1996) in Arlequin with 1000 permutations ($\alpha = 0.05$). *P* values were adjusted using the Bonferroni

correction (Rice 1989). Measures of genetic variation and descriptive statistics for two populations including number of alleles (N_A), allelic richness (*AR*), and inbreeding coefficient (F_{is}) were calculated using the software program FSTAT version 2.9.3.2 (Goudet 1995). To enable comparison of diversity across sites, allelic richness (*AR*) was calculated for each locus at both sites (in addition to number of alleles), that corrects for different sample sizes using rare faction (leberg 2002).

Table 1. Summary data for SJT (*Katsuwonus pelamis*) microsatellites loci

Locus	Repeat motiff		Primer sequence (5'-3')	Product size (bp)	Size range	T_a ^{0}C.	GenBank accession number
UTD535	$(AGAT)_9$	F	CAC TGA AGA TAT AGG CAG CCT TG	193	(150-218)	55	HM631826
		R	TTT CTC CAG CGG CAT TAC AT				
UTD523	$(GATA)_{18}$	F	TTT GAA TGG GAG ACA TGC AG	247	(172-268)	55	HM631818
		R	TGT CCT GCA CTT GTG TTC ACT				
UTD172	$(GACT)_5$	F	GTT GTG TAT TTT TGG CTG GAC C	145	(118-158)	55	HM631813
		R	CAA CAG CTA ACG GGC AAA TTC C				
UTD328	$(GCT)_8$	F	GAG AGA GAA GCG GAC AGG ATA GG	143	(120-157)	50	HM631815
		R	TGA GTA ATA GAG AGT GGG AAT GG				
UTD203	$(GAA)_7CT(GAA)_2$	F	CCC TGT GCT GTC TGT GAA G	157	(134-161)	50	HM631814
		R	TTG AAT CAA TGG CAA CTG GA				
UTD73	$(AACT)_6$	F	TGT GTG ATG AAG CTA AAG	135	(148-188)	50	HM631828
		R	CAA AAA TAT AGC CTT CGT				
UTD329	$(AACT)_7$	F	TAC TGG GTG ATG AAG CTA AAG AC	146	(136-172)	55	HM631816
		R	TCG TAA GGG AAT ATA AAA AAG TG				
UTD531	$(ATCT)_{16}$	F	GCA GTC CTG TGG GTG ATT AAA	201	(198-246)	55	HM631823
		R	GGT AAG TAT CAG AGG CTC TAC CAT C				
UTD149	$(GGA)_{11}$	F	ACC GGT GGC TTG AAG ATT GAC AG	262	na	56	HM631812
		R	GTA AAG CTC TCT CTC CTC TCC CT				
UTD 522	$(GATA)_{17}$	F	GATTATGTTCAGTGTTCCAAGCTC	389	na	58	HM631817
		R	CACAGACAGGAAAGCAATCA				
UTD526	$(GATA)_{28}$	F	GCT CTA AAT TAA ATG GAG CAT CAA A	245	na	52	HM631819
		R	GCA GAA TCC AGT CTA GTG CAA A				
UTD528	$(CTAT)_{11}$	F	GGC CTA GCT AGC AGA ATC ACT C	150	na	54	HM631820
		R	AGT GCC ATT GAA CCC ACC TA				

UTD529	$(GACA)_4$ GACGA $(ATAG)_{22}$	F	ACCCAGCAATTGACATCTGA	245	na	58	HM631821
		R	ACTAATGAATTCGCGGCC				
UTD530	$(TAGA)_{14}$ TATA $(TAGA)_5$	F	GTT TAA GGC CTA GCT AGC AGA A	188	na	52	HM631822
		R	TCC CCG AGA GTG AAA ATG TC				
UTD532	$(TATC)_{21}$	F	GGC CTA GCT AGC AGA ATC CA	190	na	52	HM631824
		R	TGC TGC CAT TAT ACC TGC AT				
UTD533	$(CTAT)_{12}$	F	ACGCGTCAGACTGCACTTC	225	na	60	HM631825
		R	GCACATATTACGGTAAATACACCG				
UTD540	$(ATAG)_{17}$	F	TCA TCC TCT CCA TTG AAC CTC	236	na	53	HM631827
		R	GGC CTA GCT AGC AGA ATC ACA				

T_a ^{0}C - Annealing temperature, F – Forward primer, R – Revers primer

Results

Measures of genetic variation and descriptive statistics for two populations are summarised in Table 2. All loci that amplified successfully were polymorphic possessing 5 to 27 alleles, respectively. Four loci (UTD172, UTD328, UTD523 and UTD531) of site MD showed significant deviations from HWE after Bonferroni correction (Rice, 1989) for multiple tests. At the same time, three of the four loci showing HWE deviations (except UTD 535) indicated a possibility of null alleles being present due to excess of homozygotes, after analysis with MICROCHECKER version 2.2.3 software (Oosterhout et al. 2004). In addition, locus 328 at site KM also showed possibility of null alleles due to excess of homozygotes. Linkage disequilibrium was not detected between any pairs of loci after Bonferroni correction.

Table 2 Descriptive statistics for two SJT (*Katsuwonus pelamis*) populations for eight microsatellite loci. Significant probability values after Bonferroni correction ($\alpha = 0.05/16 = 0.0031$)

		Locus								
Population		UTD535	UTD523	UTD172	UTD328	UTD203	UTD73	UTD329	UTD531	Avg. across loci
KM	N	37	33	34	43	44	40	31	25	35.875
	N_A	14	19	7	9	5	8	7	12	10.125
	AR	13.171	17.434	6.65	8.376	4.059	6.825	6.418	12	9.367
	H_E	0.914	0.925	0.676	0.820	0.471	0.688	0.625	0.918	0.755

	*F*is	0.115	0.051	-0.356	0.179	-0.308	0.02	0.072	0.131	0.016
MD	N	41	43	45	51	53	51	48	39	46.375
	N_A	16	27	10	12	7	11	11	14	13.500
	AR	14.139	23.131	8.824	10.365	5.138	9.012	9.481	12.772	11.608
	H_E	0.914	0.961	0.827	0.857	0.450	0.720	0.760	0.907	0.800
	*F*is	0.175	0.422*	-0.131*	0.224*	-0.006	0.047	0.124	0.551*	0.200

N-sample size, N_A-number of alleles, *AR*-allelic richness, H_E-expected heterozygosity, *F*is-inbreeding coefficient, *- significant deviations from the HWE

Discussion

The single tri- and seven tetra-nucleotide microsatellite markers optimised here show a high level of polymorphism, according to number of alleles and allelic richness at each locus and at both sites (Table 2). The most important and the most beneficial factor of these tri- and tetra-nucleotide microsatellite markers compared to di-nucleotides, is minimum level of stutter bands in tri- and tetra- nucleotide microsatellite markers. Further, stutter bands in di-nucleotide microsatellite markers can easily and frequently cause scoring errors once the two alleles are closely spaced (Perlin et al. 1995).

Six of these eight microsatellite markers have used successfully screening over 333 individuals from nine sampling sites around Sri Lanka, the Laccadive Islands, and the Maldive Islands to characterise population structure and extent of natural gene flow in SJT populations from the north western Indian Ocean (Dammannagoda et al. 2011). In addition, these markers are being used currently to screen SJT populations from the eastern Indian Ocean around Malaysia (Bhassu, S. pers.comm.).

In this analysis, deviations from HWE at some loci had been expected however, as the populations screened here were comprised of individuals from two divergent SJT mtDNA clades (Dammannagoda et al. 2011), potentially causing a Wahlund effect as indicated by excess of homozygotes at these loci (Chand et al. 2005). Of the total 17 primer pairs developed in this study,

only 8 could optimise to a condition that is suitable for large scale screening of individuals, due to time and fund restrictions. The seven tetranucleotide and the single trinucleotide polymorphic microsatellite loci described here can provide a powerful addition to the markers already available for SJT and can contribute to development of a better understanding of the scale of individual dispersal, population structure within and between wild SJT populations, effective population sizes of each clade for the conservation and management of wild stocks, and for investigating evolutionary processes underlying genetic divergence among populations across major oceans around the world.

References

REFERENCES

Broughton, R.E. and Gold, J.R. (1997). Microsatellite development and survey of variation in northern bluefin tuna (*Thunnus thynnus*). *Molecular Marine Biology and Biotechnology* **6**:308-314.

Chand,V., De Bruyn, M. and Mather, P.B. (2005). Microsatellite loci in the eastern form of the giant freshwater prawn (*Macrobrachium rosenbergii*). *Molecular Ecology Notes* **5**:308-310.

Clark, T.B., Ma, L., Saillant, E., and Gold, J.R. (2004). Microsatellite DNA markers for population-genetic studies of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) and other species of genus *Thunnus*. *Molecular Ecology Notes* **4**:70-74.

Dammannagoda, S.T., Hurwood, D.A. and Mather, P.B. (2011). Genetic analysis reveals two stocks of skipjack tuna (*Katsuwonus pelamis*) in the north western Indian Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* **68(2)**: 210-223.

Ely, B., Vinas, J., Alvarado Bremer, J., Black, B., Lucas, L., Covello, K., Labrie, A.V. and Thelen, E. (2005). Consequences of the historical demography in the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (*Thunnus albacares*) and skipjack tuna (*Katsuwonus pelamis*). *BMC Evolutionary Biology* **5**:19.

FAO (2011). FAO fishery statistical collections: Global tuna nominal catches-online query. http://www.fao.org/fishery/statistics/tuna-catches. Accessed 20 Feb 2011.

Glaubitz, J.C. (2004). CONVERT: A user-friendly program to reformat diploid genetic data for commonly use population genetics software packages. *Molecular Ecology Notes* **4**:309-310.

Goudet, J. (1995). FSTAT (Version 1.2): a computer program to calculate F-statistics. *Heredity* **86**: 485-486.

Graves J.E., Ferris S.D. and Dizon, A.E. (1984). High genetic similarity of Atlantic and Pacific skipjack tuna demonstrated with restriction endonuclease analysis of mitochondrial DNA. *Marine Biology* **79**:315-319.

Guo, S. and Thompson, E. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. **48**:361-372.

Leberg, P.L. (2002). Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology* **11**: 2445-2447.

McDowell, J.R., Diaz-Jaimes, P. and Graves, J.E. (2002). Isolation and characterization of seven tetranucleotide microsatellite loci from Atlantic northern bluefin tuna *Thunnus thynnus thynnus*. *Molecular Ecology Notes* **2:** 214-216.

Menezes, M.R., Ikeda, M. and Taniguchi, N. (2005). Genetic variation in skipjack tuna *Katsuwonus pelamis* (L.) using PCR-RFLP analysis of the mtDNA D-loop region. *Journal of Fish Biology* **68**: Supplement A, 156-161.

Menezes, M.R., Noguchi, D., Nakajima, M. and Taniguchi, M. (2008). Microsatellite development and survey of genetic variation in skipjack tuna *Katsuwonus pelamis*. *Journal of Fish Biology* **73**:463-473.

Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from nucleated cells. *Nucleic Acid Research* **16**:1215.

Oosterhout, C., Hutchinson, W.F., Wills, P.M.D. and Shipley, P. (2004). Microchecker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**: 535-538.

Perlin, M.W., Lancia, G. and Ng, S.K. (1995). Towards fully automated genotyping: genotyping microsatellite markers by deconvolution. *American Journal of Human Genetics* **57**:1199-2210.

Rice, W.J. (1989). Analysing tables of statistical tests. *Evolution* **43**: 223-225.

Rozen, S. and Skaetsky, H. (2000). Primer3 on the WWW for General Users and for Biologist Programmers. In: Stephen, M. and Stephen A. K. (Eds) *Bioinformatics Methods and Protocols,* Humana Press, MA, UK 132: 365-386.

Schneider, S., Roesslli, D. and Excoffier, L. (2005). Arlerquin version 2.00: a software for population genetics data analysis.

Slatkin, M. and Excoffier, L. (1996). Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. *Heredity*. **76**: 377-383.

Takagi, M., Okamura, T., Chow, S. and Taniguch, N. (1999). PCR primers for microsatellite loci in tuna species of the genus *Thunnus* and its application for population genetic study. *Fisheries Science* **65**:571-576.

Toth, G., T., Zoltan G., and Jerzy J. (2000). Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Research* **10**: 967-981.