A peer-reviewed version of this preprint was published in PeerJ on 30 August 2016.

View the peer-reviewed version (peerj.com/articles/2205), which is the preferred citable publication unless you specifically need to cite this preprint.

Mendes NJ, Cruz VP, Ashikaga FY, Camargo SM, Oliveira C, Piercy AN, Burgess GH, Coelho R, Santos MN, Mendonça FF, Foresti F. 2016. Microsatellite loci in the tiger shark and cross-species amplification using pyrosequencing technology. PeerJ 4:e2205 https://doi.org/10.7717/peerj.2205
Microsatellite loci in the tiger shark and cross-species amplification using pyrosequencing technology

Natália J Mendes, Vanessa P Cruz, Fernando Y Ashikaga, Sâmia M Camargo, Claudio Oliveira, Andrew N Piercy, George H Burgess, Rui Coelho, Miguel N Santos, Fernando F Mendonça, Fausto Foresti

The tiger shark (*Galeocerdo cuvier*) has a global distribution in tropical and warm temperate seas, and is caught in numerous fisheries worldwide, mainly as bycatch. It is currently assessed as near threatened by the International Union for Conservation of Nature (IUCN) Red List. In this study we identified 9 microsatellite loci through next generation sequencing (454 pyrosequencing) using 29 samples from the western Atlantic. The genetic diversity of these loci was assessed and revealed a total of 48 alleles ranging from 3 to 7 alleles per locus (average of 5.3 alleles). Cross-species amplification was assessed in three other species: *Carcharhinus longimanus*, *C. acronotus* and *Alopias superciliosus*. Given the potential applicability of genetic markers for biological conservation, these data may contribute to the population assessment of this and other species of sharks worldwide.
Microsatellite loci in the tiger shark and cross-species amplification using pyrosequencing technology

Mendes NJ¹, Cruz VP¹, Ashikaga FY¹, Camargo SM ¹–⁵, Oliveira C¹, Piercy A², Burgess G³, Coelho R⁴, Santos MN⁴, Mendonça FF⁵, Foresti F¹

¹Laboratório de Biologia e Genética de Peixes, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil.
²Lake Nona Campus, Valencia College, Florida, USA.
³Florida Program for Shark Research, Florida Museum of Natural History, University of Florida, Florida, USA.
⁴Portuguese Institute for the Ocean and Atmosphere, IPMA, Algarve, Portugal.
⁵Laboratório de Genética Pesqueira e Conservação, Marine Sciences Institute, Universidade Federal de São Paulo, Santos, São Paulo, Brazil.

Corresponding Author: Fernando Fernandes Mendonça
Universidade Federal de São Paulo – UNIFESP, Instituto do Mar, Laboratório de Genética Pesqueira e Conservação, Av. Almirante Saldanha da Gama, 89. Ponta da Praia, Santos, SP – Brasil - CEP 11030-400 e-mail: fernando.mendonca@unifesp.br

Running title: Microsatellite in the tiger shark by pyrosequencing
Abstract

The tiger shark (*Galeocerdo cuvier*) has a global distribution in tropical and warm temperate seas, and is caught in numerous fisheries worldwide, mainly as bycatch. It is currently assessed as near threatened by the International Union for Conservation of Nature (IUCN) Red List. In this study we identified 9 microsatellite loci through next generation sequencing (454 pyrosequencing) using 29 samples from the western Atlantic. The genetic diversity of these loci was assessed and revealed a total of 48 alleles ranging from 3 to 7 alleles per locus (average of 5.3 alleles). Cross-species amplification was assessed in three other species: *Carcharhinus longimanus*, *C. acronotus* and *Alopias superciliosus*. Given the potential applicability of genetic markers for biological conservation, these data may contribute to the population assessment of this and other species of sharks worldwide.
Introduction

The tiger shark *Galeocerdo cuvier* (Péron & Lesueur 1822), is a shark from the order Carcharhiniformes and family Carcharhinidae. It has a worldwide distribution in tropical and temperate seas, and is considered a top predator generally requiring large foraging areas (Heupel et al. 2014). Recent data show that this species can move long distances and occupies different habitats, including coastal regions and is therefore more susceptible to anthropogenic threats (Heupel et al. 2014).

Captured in many world fisheries as bycatch, the tiger shark is currently classified as "Near Threatened" by the International Union for Conservation of Nature (IUCN). However, some basic information, such as the characterization of population genetic structure, identification of geographical restrictions to gene flow with possible local populations remain broadly unknown. In light of this, information on their conservation status is difficult to assess. For this reason, molecular markers have been increasingly used in species conservation and management programs, including microsatellite molecular markers (Simple Sequence Repeats - SSR). A range of SSR markers have been developed using the pyrosequencing technique, generating information with millions of base pairs in a single run and in a short period of time.

Specifically for the tiger shark, nine SSR markers were previously developed on specimens from the Hawaiian archipelago (Bernard et al. 2015), but cross-application was not tested for other shark species. Thus, the objectives of this study were to identify other microsatellites for the tiger shark in specimens from the Atlantic, and design additional molecular markers that can be used in this and other shark species for population genetics studies.
Material and Methods

Sampling

In fulfillment of data archiving guidelines (Baker 2013), primary data have been deposited with Dryad. Samples of tiger shark were collected in landings of the fishing fleet from São Paulo coast (n=12) and in scientific cruises in the Fernando de Noronha archipelago (n=6) by researchers from the Biosciences Institute of Botucatu, São Paulo State University, and Marine Sciences Institute of the São Paulo Federal University, in Brazil. Additionally, 11 samples were collected from the east coast of Florida, by the Florida Program for Shark Research, University of Florida, USA. For evaluating the cross-amplification we used 6 samples of *Carcharhinus acronotus* collected in São Paulo coast, 5 samples of *C. longimanus* and 5 samples of *Alopias superciliosus*, collected in the northeast Atlantic by onboard observers from the Portuguese Institute for the Ocean and Atmosphere (IPMA), Portugal. All sampled tissues were stored in 95% ethanol to ensure the integrity and quality of tissue for molecular analysis.

454 GS-FLX Pyrosequencing and Microsatellite Discovery

The total genomic DNA was extracted from each sample following the protocol described by Ivanova et al. (2006). A sample of 100 µg of tiger-shark DNA from São Paulo coast was sequenced on a Roche 454 GS FLX sequencer with Titanium platform “Genome sequencer 20 System” (Instituto Agrobiotecnológico de Rosário – INDEAR, Argentina), following procedures described in Margulies et al. (2005).

To isolate microsatellites and design primers for population genetics all sequences of the SSR were compiled using Primer3 (Rozen & Skaletsky 2000) and BatchPrimer3 (You et al. 2008). Primers were designed based on the following criteria: primer size of 20 bp (min = 18,
max = 22 bp), ideal annealing temperature of 60°C degrees (min = 55 °C, max = 63 °C), GC optimum of 60% (= 40% min, max = 80%) and the size of the amplified product ranging from 50-500 bp. The sequences were then grouped and aligned in the Clustal W software (Thompson et al. 1994), identifying duplicated sequences for the same locus.

Novel Microsatellite Markers

The PCR amplifications to test the synthesized primers were performed in a Thermal Cycler Veriti™ (Applied Biosystems™, Life Technologies) under the following conditions: initial denaturing for 10 min at 95 °C; 30 cycles of 94 °C for 45s, the primer annealing temperature (TA) was tested from of 51 °C to 57 °C for 50 s; 72 °C for 50s, and a final extension at 72 °C for 20min. The total reaction volume was 10 µL and composed of 0.20 X PCR Buffer, 0.25 mM MgCl₂, 0.05 mM of each dNTP, 0.5 units of Platinum Taq DNA polymerase (Invitrogen™, Life Technologies, EUA), 0.10 µM reverse primer, 0.10 µM forward primer, and 30 ng of template DNA.

To verify the effectiveness of the reaction and the amplification of the fragments, 1.5 µl of the PCR product were subjected to electrophoresis on a 1% agarose gel. The amplified products were compared with a 1Kb plus ladder (Invitrogen), subsequently visualized on a transilluminator and photographed with a digital camera using the Kodak Digital Science software.

Genotyping was done with the M13-tail PCR method of Schuelke (2000). The best loci, that showed high polymorphism and quality of bands, were selected and further analyzed on an ABI 3130xl sequencer (Applied Biosystems™, Life Technologies). The allele sizes were
determined using ROX 500 (Applied Biosystems) as an internal standard with the software package GeneMapper 3.7 (Applied Biosystems).

We used the software GenAlex analysis 6.1 (Peakall & Smouse, 2006) to convert our data to run in other analysis programs. Arlequin 3.5 (Guo & Thompson, 1992) was used to calculate heterozygosity, number of alleles, Hardy-Weinberg equilibrium and linkage disequilibrium. The program Cervus v.3.0.7 (Marshall et al. 1998) was used to test for the presence of null alleles and estimate, the inbreeding coefficient (Fis) and polymorphism information content (PIC).

Results and Discussion

From the genomic material generated by the pyrosequencing technology, a total of 71,059 reads with an average size of 367 bp was obtained, consisting of 26,075,405 nucleotides, which accounts for approximately 0.75% of the G. cuvier genome, assuming a genome size of 3.44Gb (estimated from the size of Rhincodon typus, Read et al. 2015). For the identification of microsatellite sequences, the online software Batch Primer3 was used, and 615 microsatellite loci were identified. A second filtration was subsequently performed with the software Primer 3.0 which identified 159 microsatellite loci. From these, we selected 30 loci which contained the best scores of each primer pair with a size of 15 - 20bp, a GC of 40-50% and little variation in the annealing temperature in the PCR reaction. Of these, 20 pairs of primers were synthesized and tested, 9 being polymorphic with 1 trinucleotide and 8 dinucleotide (Table 1). The sequences with polymorphic microsatellite markers in this study have been deposited in GenBank (Accession numbers: KT598263-KT598271).
The application of the developed markers resulted in 48 alleles, with a minimum of 3
(TIG_25) to 7 (TIG_1, TIG_7, TIG_12) and average of 5.3 alleles per loci. Transferability tests
of the markers in other species showed positive amplification in *C. longimanus, Alopias
superciliosus* and *C. acronotus*. For the *C. acronotus*, two loci were polymorphic (TIG_17,
TIG_5), and for *C. longimanus* and *A. superciliosus* only one polymorphic locus were observed
in 5 samples of each species, TIG_15 and TIG_7 respectively (Table 2).

In tiger shark the observed heterozygosity (Ho) and expected heterozygosity (He) ranged
from 0.16 (TIG_17) to 1.0 (TIG_10) and 0.20 (TIG_25) to 0.72 (TIG_7) respectively. The Ho
was higher than He, suggesting an excess of heterozygotes relative to the model of Hardy-
Weinberg equilibrium (HWE). Significant differences from Hardy-Weinberg equilibrium after
Bonferroni correction (p<0.01) were detected in only 2 loci (TIG_10 and TIG_17). The deviation
in the Hardy-Weinberg equilibrium for locus TIG_17 (0.715) can be explained by a significant
value in intrapopulation inbreeding coefficient (Fis) (Kordicheva et al. 2010). This locus was the
only one with a positive value for the Fis, and may be evidence of a heterozygous deficiency
(Holsinger & Weir 2009), resulting in a decrease in genetic variability.

Imbalance values in Hardy-Weinberg equations when considering microsatellite locus
may be due to the presence of null alleles (Kordicheva et al. 2010). However, the presence of
null allele was not detected in the present study, indicating that the markers developed are of
high quality. Further, the polymorphism information content (PIC) was highly informative for all
the loci (PIC > 0.5), also indicating a high quality marker (Botstein et al. 1980).

In the present study, the average expected heterozygosity was approximately 0.50 and the
average observed heterozygosity was 0.55. The levels of genetic variability seen in this study
may be due to population differences resulting from remote sample locations. This is to be
expected given that the samples are coming from different oceans and the finding of significant
differences in the levels of heterozygosity among different groups would not be unforeseen.

**Supplementary Material**

Supplementary material can be found at http://www.ncbi.nlm.nih.gov/genbank/

**Acknowledgements**

The authors thank all those who contributed to this study, including the researchers who
carried out the sample collection in Fernando de Noronha archipelago and the fishermen for
facilitating tissue collection during landings in the São Paulo coast. Sampling and data collection
from the Portuguese fishery were obtained and supported by the Programa Nacional de
Amostragem Biologica (PNAB) within the scope of the EU Data Collection Framework (DCF).
The authors are grateful to all fishery observers and longline skippers that helped collect samples
for this study.

**References**

Baker CS. 2013. Journal of Heredity Adopts Joint Data Archiving Policy. J Hered 104:1. doi:
10.1093/jhered/ess137

Bernard AM, Feldheim KA, Shivji MS. 2015. Isolation and characterization of polymorphic
microsatellite markers from a globally distributed marine apex predator, the tiger shark
(Galeocerdo cuvier). Conserv Genet Resour. 7:509–511.
Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet.* 32: 314-331.

Guo X, Elston RC. 1999. Linkage information content of polymorphic genetic markers. *Hum Hered.* 49: 112-118.

Heupel M, Knip D, Simpfendorfer C, Dulvy N. 2014. Sizing up the ecological role of sharks as predators. *Mar Ecol Prog Ser.* 495: 291–298.

Holsinger KE, Weir BS. 2009 Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Rev Genet.* 10, 639–650.

Kordicheva SY, Rubtsova GA, Shitova MA, Shaikhaev GO, Afanasiev KI, Zhivotovsky LAA. 2010. Search for null alleles at the microsatellite locus of chum salmon (*Oncorhynchus keta Walbaum*). *Russ J Genet.* 46: 1019-1022.

Margulies M, Egholm M, Altman W, Attiya S, Bader J, Bemben L. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 437: 376–380.

Marshall T, Slate J, Kruuk L, Pemberton J. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol.* 7: 639–655

Peakall ROD, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes.* 6: 288-295.

Read TD, Petit III RA, Joseph SJ, Alam MT, RW, Ahmad M, Bhimani R, Vuong JS, Haase CP, Webb H, Dove ADM. 2015 Draft sequencing and assembly of the genome of the world’s largest fish, the whale shark: *Rhincodon typus* Smith 1828. PeerJ PrePrints 3:e1036 https://dx.doi.org/10.7287/peerj.preprints.837v1

Rozen S, Skaletsky H. 1999. Primer 3 on the WWW for general users and for biologist programmers. Totowa(NJ): Human Press Inc
Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotech.* 18: 233-234.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics.* 9: 253.
Table 1. General information about the microsatellite analysis.

Table 2. Data for microsatellite loci of the cross-amplification in *Carcharhinus longimanus*, *Carcharhinus acronotus* and *Alopias superciliosus*.

Table 3. Data for microsatellite loci in the tiger shark, *Galeocerdo cuvier*.
Table 1 (on next page)

General information about the microsatellite analysis.
| Analyzes                                           | N of sequences |
|---------------------------------------------------|----------------|
| Number of reads                                   | 71.059         |
| Selection of microsatellites (using BatchPrimer3) | 615            |
| Secondary selection of microsatellite (using Primer 3.0) | 159            |
| Amplification and control of PCR product on agarose gel | 30             |
| Microsatellite loci to synthesize with fluorescent dye | 20             |
| Polymorphism test with capillary sequencer         | 10             |
| Microsatellite *loci* in linkage equilibrium       | 9              |
Data for microsatellite loci of the cross-amplification in *Carcharhinus longimanus*, *Carcharhinus acronotus* and *Alopias superciliosus*. 
| Loci  | Na | Range(bp) | Loci  | Na | Range(bp) | Loci  | Na | Range(bp) |
|-------|----|-----------|-------|----|-----------|-------|----|-----------|
| TIG_1 | 2  | 116-118   |       |    |           |       |    |           |
| TIG_5 | 3  | 260-264   |       |    |           |       |    |           |
| TIG_7 | 2  | 170-180   |       |    |           |       |    |           |
| TIG_10| 2  | 251-253   |       |    |           |       |    |           |
| TIG_12| 2  | 296-364   |       |    |           |       |    |           |
| TIG_15| 1  | 336       |       |    |           |       |    |           |
| TIG_17| 3  | 242-270   |       |    |           |       |    |           |
| TIG_19| 0  | 0         |       |    |           |       |    |           |
| TIG_25| 1  | 396       |       |    |           |       |    |           |

Na: number of alleles
Table 3 (on next page)

Data for microsatellite loci in the tiger shark, *Galeocerdo cuvier*.
| LOCI  | PRIMER SEQUENCE (5'→3')              | MOTIF | T °C | N  | NA | RANGE(bp) | HO   | HE   | HWE   | FIS  | PIC  | F(NULL) |
|-------|--------------------------------------|-------|------|----|----|-----------|------|------|-------|------|------|--------|
| TIG_1 | F_CTCTTGACGTTGCTGAGTC               | (AC)10| 53   | 29 | 7  | 116-154   | 0.758| 0.642| 0.711 | -0.184| 0.710| -0.194 |
|       | R_AATGGCAACTTTTCCTGTCC             |       |      |    |    |           |      |      |       |      |      |        |
| TIG_5 | F_GACCAACCTCCCATCACG             | (CT)8 | 51   | 26 | 4  | 203-257   | 0.384| 0.337| 1.000 | -0.141| 0.589| -0.239 |
|       | R_AGAGGGAAGTGGTGTGGG               |       |      |    |    |           |      |      |       |      |      |        |
| TIG_7 | F_CACCAACCTCCCCATCAC              | (AC)15| 57   | 27 | 7  | 169-183   | 0.925| 0.726| 0.318 | -0.280| 0.811| -0.101 |
|       | R_CAGACATTCCTCCCTCCATCC           |       |      |    |    |           |      |      |       |      |      |        |
| TIG_10| F_CTCAGCAGGTCTGGACAAACA            | (GT)10| 59   | 29 | 5  | 256-276   | 1.000| 0.655| 0.000 | -0.539| 0.608| -0.245 |
|       | R_GGTGGTAGGACATGGAAACG             |       |      |    |    |           |      |      |       |      |      |        |
| TIG_12| F_TGCCCCATGAGTCGCTTTTTCT         | (CA)11| 53   | 28 | 7  | 364-376   | 0.535| 0.520| 0.543 | -0.030| 0.682| -0.213 |
|       | R_TGCCCATGTGGTACTGCTAC            |       |      |    |    |           |      |      |       |      |      |        |
| TIG_15| F_AACTGCCAAAAGGGGACAG             | (TG)15| 55   | 25 | 6  | 231-241   | 0.520| 0.463| 0.675 | -0.124| 0.650| -0.233 |
|       | R_GTAAGCCAAAAGGACAGACCTCC         |       |      |    |    |           |      |      |       |      |      |        |
| TIG_17| F_TGAAAGCTAACGACGGGTCTCTG         | (GT)11| 57   | 25 | 4  | 268-286   | 0.160| 0.554| 0.000 | 0.715 | 0.734| -0.138 |
|       | R_AGCCGAGAAAGATCAAGAGGA           |       |      |    |    |           |      |      |       |      |      |        |
| TIG_19| F_TGCTTTTGTGTGCTGAGGTAGTG        | (TG)10| 53   | 27 | 5  | 337-353   | 0.555| 0.443| 0.677 | -0.260| 0.627| -0.214 |
|       | R_TGGAAGGTTCAATCCAGGAC            |       |      |    |    |           |      |      |       |      |      |        |
| TIG_25| F_CCGTGCCCTATGTTGGATTTCT          | (CCT)5| 55   | 27 | 3  | 331-349   | 0.222| 0.206| 1.000 | -0.075| 0.511| -0.285 |
|       | R_CTTGGAAGAGTGGGCGAAC            |       |      |    |    |           |      |      |       |      |      |        |

T °C: primer annealing temperature, N: number of individuals analyzed, NA: number of alleles, He: expected heterozygosity, Ho: observed heterozygosity, HWE: probability of departure from Hardy–Weinberg equilibrium, Fis: inbreeding coefficient, PIC: polymorphism information content, F(NULL): null alleles.