

Isolation of Twenty-Five New Molecular Microsatellite Markers from *Alligator mississippiensis* (Alligatoridae, Alligatorinae) EST Sequences using in Silico Approach

Rodrigo Barban Zucoloto

Clara Ribeiro Porto

Universidade Federal da Bahia/UFBA

Instituto de Biologia

Departamento de Biologia Geral

Laboratório de Genética de Populações e Evolução Molecular

Salvador, Bahia, 40170-290

Brasil

Abstract

Microsatellite markers have been applied to conservation genetic studies of crocodilians since the second half of 90's. The identification of highly transferable markers would be very important to crocodilian genetic studies. Here is described the identification of twenty-five new microsatellite markers from Alligator mississippiensis (Daudin, 1802) EST sequences and discussed their expected efficiency for the amplification of DNA of other crocodilian species.

Keywords: Alligatorinae, STR, SSR, CID, crocodilians

1. Introduction

Conservation genetics is a research field aging about 23 yr that concentrates efforts to apply molecular genetic analysis to solve questions concerning species conservation (Ogden, Dawnay, & McEwing, 2009). Microsatellite markers have been applied to conservation genetic studies of crocodilians since the second half of 90's, including works about isolation of new microsatellite markers and cross-species amplification (Chaeychomsri, Chaeychomsri, & Tuntirungkij, 2008; Chaeychomsri & Tabthipwon, 2008; Chaeychomsri, 2008; FitzSimmons et al., 2001; Glenn, Dessauer, & Braun, 1998; Jing, Wang, Lan, & Fang, 2008; Miles, Isberg, Moran, Hagen, & Glenn, 2008; Miles, Lance, Isberg, Moran, & Glenn, 2009; Oliveira, Farias, Marioni, Campos, & Hrbek, 2010; Subalusky, Garrick, Schable, Osborne, & Glenn, 2012; Villela, Coutinho, Piña, & Verdade, 2008; Wu, Wu, & Glenn, 2012). This amount of research prove the useful of microsatellite markers in studies of crocodilian genetics. The identification of highly transferable markers will be very important to crocodilian genetic studies particularly to brazilian species whose markers were limited to those isolated in a few works (Oliveira et al., 2010; Villela et al., 2008).

The strategies for microsatellite isolation in most of these works were from genomic libraries of DNA or from the enrichment protocol (Chaeychomsri & Tabthipwon, 2008; Chaeychomsri, 2008; Glenn et al., 1998; Jing et al., 2008; Miles et al., 2008; Oliveira et al., 2010; Subalusky et al., 2012), and a combined method involving enrichment and 454 pyrosequencing (Wu et al., 2012). Here is described the identification of new microsatellite markers from *Alligator mississippiensis* EST (Expressed Sequence Tags) sequences as a strategy for microsatellite isolation and discussed their expected efficiency for the amplification of DNA of crocodilian species.

2. Material and Methods

A total of 5425 EST sequences from *Alligator mississippiensis* were retrieved from the subset of Alligatorinae subfamily EST sequences (taxonomic id 34915) of the NCBI databank. All EST sequences retrieved were certified from *Alligator mississippiensis* (taxonomic id 8496).

This subset of EST sequences was divided into 20 files with an average of 272 sequences and submitted to CID available online at (<http://www.shrimp.ufscar.br/cid/index.php>). The subdivision of original subset of EST sequences was necessary due to limitations of CID implementation.

CID is a pipeline web implementation that includes Primers 3 and other programs used to pick primers for PCR of the sequences containing microsatellites (Freitas, Martins, & Jr PM Galetti, 2008). Each file had its results presented as a worksheet with analysis of the sequences. The searching criteria for CID analysis followed the subsequent definitions for unite size per minimum number of repeats: almost ten repeats for dinucleotide motifs, four repeats for trinucleotide motifs, three repeats for tetranucleotide motifs, three repeats for pentanucleotide motifs and three repeats for hexanucleotide motifs. The maximum number of bases interrupting two microsatellites in a compound microsatellite was defined as 100 bp (base pairs). After CID analysis it was noted that some sequences were redundant because some microsatellites identified in different sequences had identical primer pairs. In these cases it indicates that the original sequences were redundant which was confirmed by sequence comparison. Assuming that the number of repeats is indicative of greater probability of polymorphism it were chosen the microsatellite markers that had the highest number of repeats. After elimination of the redundancy and the breakdown according to the number of repeats, the primers for the chosen microsatellite markers were submitted to inspection of harpin and primer-dimmer with the software's Oligo Analyser 1.5, Oligo Explorer 1.5 and Sequence Manipulation Suite 2.0.

3. Results

From a total of 5425 EST sequences of *Alligator mississippiensis* 237 sequences presented microsatellites, however after excluding the observed redundancy the number of EST sequences presenting microsatellite was 153. It were identified seven microsatellite markers with compound motifs from which three were chosen: *Alliμ1*, *Alliμ2* and *Alliμ3*; 14 with dinucleotide motifs: *Alliμ12* to *Alliμ25*; 55 with trinucleotide motifs from which four were chosen: *Alliμ4*, *Alliμ5*, *Alliμ6* and *Alliμ7*; 41 with tetranucleotide motifs from which one were chosen: *Alliμ8*; 31 with pentanucleotide motifs from which two were chosen: *Alliμ9* and *Alliμ10*; and five hexanucleotide motifs from which one were chosen, *Alliμ11*. The results were shown in Table 1.

T1

4. Discussion

The observation that EST sequences contain a higher proportion of microsatellite markers has been reported (Durand et al., 2010; Kim, Ratcliffe, French, Liu, & Sappington, 2008; Pérez et al., 2005; Rohrer, Fahrenkrug, Nonneman, Tao, & Warren, 2002). These authors shown that EST based microsatellite gave acceptable level of polymorphism for genetic studies.

Among the porcine EST-microsatellites isolated by Rohrer et al., (2002) dinucleotide repeat markers are more polymorphic than tri to hexanucleotide repeat markers tested, it was observed that 72% of dinucleotide markers were informative relatively to only 7% of other repeat motifs. Pérez et al., (2005) noted that 69% of EST-microsatellite were transferable within the genus *Litopenaeus*. According to Kim et al., (2008) comparing noncoding, nontranscribed regions with EST- microsatellites, it is observed that generally EST based microsatellites are less polymorphic but had reduced occurrence of null alleles and greater cross-species amplification. According to Durand et al., (2010) the polymorphism of microsatellites isolated from EST sequences is low than the genomic isolated ones, but the transferability to chestnut, a phylogenetically related species to oak was higher.

In agreement with the results presented above from a variety of species, the twenty-five new molecular microsatellite markers isolated from *Alligator mississippiensis* EST sequences should be enough polymorphic and highly transferable among crocodylian species, specially those from Alligatorinae subfamily, which could be great important to its conservation genetic studies.

Acknowledgments

I would like to thank Dra. Flora Maria de Campos Fernandes for her help with the manuscript and students Gabriel Vieira and Rafaela Marocci to collect partial data. Partial financial support came from Santander Bank.

References

- Chaeychomsri, W. (2008). Development of molecular markers for identification of Thai Crocodiles. Kasetsart University.
- Chaeychomsri, W., Chaeychomsri, S., & Tuntirungkij, M. (2008). Characterization of Microsatellite Markers for the Siamese Crocodile and Amplification in the Closely Related Genus *Crocodylus*. *Kasetsart J. (Nat. Sci.)*, 42, 682–692.
- Chaeychomsri, W., & Tabthipwon, P. (2008). Development of Microsatellite Markers for Siamese Crocodile (*Crocodylus siamensis*). *Kasetsart J. (Nat. Sci.)*, 42, 256–262.
- Durand, J., Bodénès, C., Chancerel, E., Frigerio, J.-M., Vendramin, G., Sebastiani, F., ... Plomion, C. (2010). A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics*, 11(570), 1–13. <http://doi.org/10.1186/1471-2164-11-570>
- FitzSimmons, N., Tanksley, S., Forstner, M., Louis, E., Daghli, R., Gratten, J., & S Davis. (2001). Microsatellite markers for *Crocodylus*: new genetic tools for population genetics, mating system studies and forensics. In G. Grigg, F. Seebacher, & C. Franklin (Eds.), *Crocodylian biology and evolution* (pp. 51–57). Australia: Surrey Beatty: Chipping Norton.
- Freitas, P., Martins, D., & Jr PM Galetti. (2008). CID (Clipping vectors, Identifying SSR, Design of primers): a rapid and efficient bioinformatic tool for the detection of simple sequence repeats (SSRs) from genomic libraries. *Molecular Ecology Notes*, 68, 107–108.
- Glenn, T. C., Dessauer, H. C., & Braun, M. J. (1998). Characterization of Microsatellite DNA Loci in American Alligators. *Copeia*, 1998(3), 591–601.
- Jing, W., Wang, X.-L., Lan, H., & Fang, S.-G. (2008). Eleven novel microsatellite markers for the Chinese alligator (*Alligator sinensis*). *Conservation Genetics*, 10(3), 543–546. <http://doi.org/10.1007/s10592-008-9561-x>
- Kim, K. S., Ratcliffe, S. T., French, B. W., Liu, L., & Sappington, T. W. (2008). Utility of EST-derived SSRs as population genetics markers in a beetle. *Journal of Heredity*, 99(2), 112–24. <http://doi.org/10.1093/jhered/esm104>
- Miles, L. G., Isberg, S. R., Moran, C., Hagen, C., & Glenn, T. C. (2008). 253 Novel polymorphic microsatellites for the saltwater crocodile (*Crocodylus porosus*). *Conservation Genetics*, 10(4), 963–980. <http://doi.org/10.1007/s10592-008-9600-7>
- Miles, L. G., Lance, S. L., Isberg, S. R., Moran, C., & Glenn, T. C. (2009). Cross-species amplification of microsatellites in crocodylians: assessment and applications for the future. *Conservation Genetics*, 10, 935–954.
- Ogden, R., Dawnay, N., & McEwing, R. (2009). Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. *Endangered Species Research*, 9, 179–195. <http://doi.org/10.3354/esr00144>
- Oliveira, D. P., Farias, I. P., Marioni, B., Campos, Z., & Hrbek, T. (2010). Microsatellite markers for mating system and population analyses of the spectacled caiman *Caiman crocodilus* (Linnaeus 1758). *Conservation Genetics Resources*, 2(1), 181–184. <http://doi.org/10.1007/s12686-010-9221-6>
- Pérez, F., Ortiz, J., Zhinaula, M., Gonzabay, C., Calderón, J., & Volckaert, F. A. M. J. (2005). Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus birdy*. *Marine Biotechnology (New York, N.Y.)*, 7(5), 554–69. <http://doi.org/10.1007/s10126-004-5099-1>
- Rohrer, G. A., Fahrenkrug, S. C., Nonneman, D., Tao, N., & Warren, W. C. (2002). Mapping microsatellite markers identified in porcine. *Animal Genetics*, 33, 372–376.
- Subalusky, A. L., Garrick, R. C., Schable, N. A., Osborne, J., & Glenn, T. C. (2012). Development and characterization of tetranucleotide microsatellite loci for the American alligator (*Alligator mississippiensis*). *Conservation Genetics Resources*, (2009). <http://doi.org/10.1007/s12686-011-9593-2>
- Villela, P. M. S., Coutinho, L. L., Piña, C. I., & Verdade, L. M. (2008). Macrogeographic genetic variation in broad-snouted caiman (*Caiman latirostris*). *Journal of Experimental Zoology. Part A, Ecological Genetics and Physiology*, 309(10), 628–36. <http://doi.org/10.1002/jez.489>
- Wu, Q., Wu, X., & Glenn, T. (2012). Fourteen novel microsatellite loci in the Chinese alligator (*Alligator sinensis*) isolated via 454 pyrosequencing. *Conservation Genetics Resources*, 4(3), 4–7. <http://doi.org/10.1007/s12686-012-9632-7>

Table 1: Microsatellite Markers

Name	GI	SSR	SSR size	SSR start	SSR end	PRIMERS (5'-3') Forward and reverse	Tm (°C)	Primer size	Size (bp)	Start (bp)	End (bp)
<i>Allip1</i>	148489706	(AGAA) ₃ (AAG) ₄	19	292	310	GAGATGGTCCAAACCAGATA GAGACGATCTGTTCTTCCAG	55 55	20 20	245	127	371
<i>Allip2</i>	148488043	(CAAG) ₃ (AGC) ₄	22	145	166	CTCCTTCCCAGTCTGACATT ACCTTCAGGATGTCACAGTTC	57 57	20 21	219	56	274
<i>Allip3</i>	148487241	(CAAG) ₃ (AGC) ₄	22	417	438	GAGTTGGGAGCTGACACTAC CACAGATGAAAGCAATGAAC	55 54	20 20	211	253	463
<i>Allip4</i>	148485897	(ATA) ₈	24	263	286	GGGGGTTTTTAATCAGAGAA CATTCCTGATAATCTGCTGG	56 56	20 20	190	164	353
<i>Allip5</i>	148488813	(GCC) ₇	21	565	585	CAAAGAGAGAGGCACACAGG ACATGTTGAGCCCGTACTTG	59 59	20 20	182	518	699
<i>Allip6</i>	148488857	(TTA) ₁₀	30	69	98	ACAAATCACTCTCTCCCCTT TAAGACCAAGATTCACCAGG	55 55	20 20	180	17	196
<i>Allip7</i>	148489394	(TTA) ₇	21	196	216	CAATACAACGCACAAATCAC TGAGAAAAGAGAGGCAAAAAG	55 55	20 20	231	133	363
<i>Allip8</i>	148487142	(TGTC) ₆	24	490	513	AAGTGAAAGCCATCAAGAAG CCTTTTTATTGGAGTCATGC	55 55	20 20	255	291	545
<i>Allip9</i>	148486702	(AAAAT) ₆	30	52	81	CCCCTGAATATGAAGTCTCTC CTGCTGCATTTCTTCTCTC	55 55	21 20	226	14	239
<i>Allip10</i>	148485752	(CAGCC) ₅	25	174	198	TCGCTGAACAGAGAACATAA GACACATCCTACTTGGCTTC	55 55	20 20	278	104	381
<i>Allip11</i>	148486297	(CTGCCT) ₅	30	174	203	TACACACCGAAGAGCAGCTT CAGTTTCCCAAGGAGCTGAG	59 60	20 20	154	127	280
<i>Allip12</i>	148485633	(AC) ₁₀	20	331	350	ACTGTGTGTGTTCTTAGGGG GCTAGGAAGGAATAGGTGGT	55 55	20 20	234	178	411
<i>Allip13</i>	148487337	(AC) ₁₃	26	468	493	CTTCCAGTCTCCACAGTA AGAAGCGACCTAAATTTCC	55 55	20 20	190	367	556
<i>Allip14</i>	148487962	(AC) ₂₁	42	370	411	CACGCTACCTTGTTGTGTAG CAAACCTGTGACATGGGTA	54 54	20 20	119	319	437
<i>Allip15</i>	148490071	(AT) ₁₂	24	366	389	CGATAACTGACCAGATGGAT GATTCAGCACAAGAACACA	55 55	20 20	277	142	418
<i>Allip16</i>	148487739	(CA) ₁₁	22	253	274	ATGCACTCATACACAGCCAG TATTAGCACAGTGATTGGCG	57 57	20 20	277	167	443
<i>Allip17</i>	148488143	(CA) ₁₂	24	95	118	CACAAGTTCTGATCCCATT AGAAGCTGAGTGGTTTCTGA	55 55	20 20	169	70	238
<i>Allip18</i>	148487745	(CA) ₁₂	24	443	466	TACCCACACACATGTACACC TTAGAATAGGAGCCTCGTG	55 55	20 20	253	390	642
<i>Allip19</i>	148486884	(CT) ₁₃	26	334	359	ACAGTTGTGAGCAACAATCA ATCGTGACGTGATGGATAAT	55 55	20 20	188	297	484
<i>Allip20</i>	148490044	(GA) ₁₁	22	383	404	GAGATGGGAAAGAGGATTTT GTCACCACACAGAGAAGGAT	55 55	20 20	222	283	504
<i>Allip21</i>	148485974	(GT) ₁₂	24	204	227	ACACTTGCCTTTAGTTGGAA ATGGGTCAAAGACAACCTGAC	55 55	20 20	146	151	296
<i>Allip22</i>	148486100	(TA) ₁₅	30	204	233	TGTGTTACAAGCCATTTGTG AGATGAATCCACCATTGCTA	56 56	20 20	139	152	290
<i>Allip23</i>	148486817	(TC) ₁₅	30	73	102	GGAAAGAGAATTTCACTGAGG TGTTCTAGGTTGATGCAAT	56 56	21 20	114	32	145
<i>Allip24</i>	148486858	(TC) ₁₅	30	63	92	GCTGTAGTCAAGCTGGAAAG CAGGAGGAAGGGAAACTATT	55 55	20 20	263	7	269
<i>Allip25</i>	148486935	(TG) ₂₅	50	260	309	TCTCTCTCTCCTCCTGGACT AGACGGTGACACAATGTCTT	56 55	20 20	160	190	349