ANALYSIS OF MICROSATELLITE MARKERS IN A CUBAN WATER BUFFALO BREED

Short Title: MICROSATELLITES IN CUBAN BUFFALOES

Authors: Odalys Uffo[1]*, Nadia Martínez[1], Atzel Acosta[1], Arianne Sanz[2], Inmaculada Martín-Burriel[2], Rosario Osta[2], Clementina Rodellar[2], Pilar Zaragoza[2]

[1] Laboratorio de Genética Molecular, Centro Nacional de Sanidad Agropecuaria, Apartado 10, PO 32700, San José de las Lajas, Mayabeque, Cuba,

[2] LAGENBIO, Facultad de Veterinaria, IA2, Universidad de Zaragoza, Miguel Servet 177, 50013, Zaragoza, España

*email : uffo@censa.edu.cu

ABSTRACT

The aim of this Regional Research Communication was to validate a panel of 30 microsatellite recommended by FAO/ISAG for studies of biodiversity in cattle to improve the characterization of Cuban buffalo populations. The water buffalo (*Bubalus bubalis*) is an economically important livestock species. Therefore, research focused on the study of the genetic relationships among water buffalo populations is useful to support conservation decisions and to design breeding schemes. Twenty-eight of the 30 tested regions were amplified, one of which (ETH10) turned out to be monomorphic. A total of 143 alleles were observed in the Cuban water buffalo population. The average number of alleles per locus was 5.04. The number of alleles per polymorphic locus ranged from two (INRA 63 and MM12) to nine (ETH185). The observed and expected heterozygosity ranged from 0.108 (HAUT24) to 0.851 (CSSM66) and 0.104 (MM12) to 0.829 (INRA32), respectively. The polymorphic information content (PIC) ranged from 0.097 (MM12) to 0.806 (INRA32), and the overall value for these markers was 0.482. Within the population, inbreeding estimates (F_{IS}) was positive in 14 of the 30 loci analyzed. This study thus highlights the usefulness of heterologous bovine microsatellite markers to assess the genetic variability in Cuban water buffalo breeds. Furthermore, the results can be utilized for future breeding strategies and conservation.

Keywords: microsatellite, STR, HWE, diversity, Cuban buffalo

INTRODUCTION

Buffalo is widely chosen for production of milk in many countries worldwide, especially in the Asian ones. In general, buffaloes produce about 1,000 kg milk/lactation, on one milking per day, and lactation length range between 250-270 days. They have shown significant response to the selection process, due to their high genetic variability with a great ability of adaptation to various environments, high fertility and the longevity in production, which allowed the herd to evolve (Vieira et al., 2011). The water buffalo has prime importance in the lives of farmers and thus in the economy of many countries worldwide. They are not only draught animals, but also a source of meat, horns, skin and particularly rich and precious milk that may be converted into cream, butter, yoghurt and many different kinds of cheese (Michelizzi et al., 2010).

The use of highly polymorphic markers as microsatellites is extremely important to investigate the genetic status. These markers are particularly important to access intra-racial diversity, levels of inbreeding, genetic differentiation between breeds, mixed breeds introgression and they are essentials in studies of genetic diversity and species conservation for purposes of selection of traits of economic interest (Freeman et al., 2005).

The water buffalo was imported into Cuba from Australia and Trinidad and Tobago Island. A total of 2,984 buffalos (2,705 swamp buffalo and 279 river buffalo) were imported in order to contribute to the agricultural economy and food security of Cuba. Current flocks far exceed the number of animals imported which is indicative of their ability to adapt to the environmental conditions in the country. Animals of river breed "Buffalypso" came from Panama and Trinidad and Tobago (Mitat, 2009). In Cuba, this population was uncontrolled mixed with swamp buffalo (Carabao) subsequently imported from Australia (Borghese and Mazzi, 2005). Recently, semen from the Mediterranean breed was imported with to improve dairy production indicators and reduce inbreeding, although there is little information about it.

The aim of this study was to validate a panel of 30 microsatellite recommended by FAO/ISAG for studies of biodiversity in cattle to improve the characterization of Cuban buffalo populations.

MATERIALS AND METHODS

Available in supplementary files

RESULTS AND DISCUSSION

The investigation of the genetic relationships among buffalo populations will provide a useful tool to support conservation decisions and to contribute to the selection and preservation of genetic resources (Attia et al., 2014). Several authors have been investigating the genetic diversity of buffalo populations worldwide using microsatellite markers (Nagarajan et al., 2009¸ Abou-Bakr et al., 2012; Acosta et al., 2014; Martínez et al., 2015).

After PCR and capillary electrophoresis, twenty-eight of the 30 tested regions were amplified successfully and only one of these markers (ETH10) turned out to be monomorphic. Using multiplex PCR amplification, microsatellite displayed reproducible and non-ambiguous peak for allele assignation. The alleles varied in their size from 69-281bp. There were observed 143 alleles in the Cuban water buffalo population, a higher than any other previously observed by Acosta et al., (2014) found a total of 87 alleles in the studied Cuban water buffalo population in their work using 16 bovine microsatellite markers. The average number of alleles per locus was 5.04, similar to the 5.44 found by Acosta el at., (2014). The number of alleles per polymorphic locus ranged from two (INRA63 and MM12) to nine (ETH185) while Acosta et al., (2014) reported values ranged from two (TGLA126) to nine (ETH225)(Table 1 supplementary file). In a different report, Martínez et al., (2015) found a total of 88 alleles across the five studied microsatellite loci when compared three different populations, one of them being Cuban. The total number of alleles per population ranged from 26 in Brazilian Murrah buffaloes to 33 in Cuban Buffalypso/Carabao hybrids. Locus CSSM033 showed the highest Na per locus (12) while ILSTS5 showed the lowest (2). Mean Na values ranged from 5.2 in Murrah to 6.6 in Buffalypso/Carabao hybrids.

The observed and expected heterozygosity ranged from 0.108 (HAUT24) to 0.851 (CSSM66), and 0.104 to 0.829 in MM12 and INRA32 loci, respectively in the studied population and were similar to reported values by Acosta et al., (2014). The polymorphic information content (PIC) (table 1) ranged from 0.097 (MM12) to 0.806 (INRA32), and the overall value for these markers was 0.482. In this work, 27 (96.4%) of the 28 bovine markers amplified were polymorphic, which confirms the conservation of DNA sequences flanking microsatellites within the *Bovidae* family. A total of 571 microsatellite markers had been characterized for water buffalo until now (Nagarajan et al., 2009); Abou-Bakr et al., (2012) analyzing a total of 471 unrelated Egyptian buffaloes found 82% of the studied markers polymorphic (9 of 11). Martínez et al., (2015) described three of the markers (ILSTS5, ETH152 and CSSM042) with PIC values below 0.5, being them moderately informative (0.5 > PIC > 0.25). Genetic markers with PIC values lower than 0.25 are considered to be less informative and those with values higher than 0.5 are reckoned as distinctly informative in population genetic studies; loci with many alleles and a PIC near one are most desirable. Following this criteria, in this study, twenty-six microsatellite loci appeared to be highly informative (PIC>0.5) and thus will be useful to evaluate the genetic diversity in Cuban buffalo population.

Fourteen loci displayed significant reduction of heterozygosity (ETH3, BM1818, ETH152, ILSTS006, INRA05, HAUT24, HEL5, INRA35, HEL9, ETH185, INRA37, INRA32, HEL1 e INRA63) (Table 1). All of loci within significant population inbreeding (F_{IS}) were out of Hardy-Weinberg equilibrium (HWE) conditions with a significant heterozygote deficit. The statistics F_{IS} is an estimate of variation within a population that measures the reduction in heterozygosity in an individual due to nonrandom mating within sub populations. The F_{IS} in Cuban water buffalo can be considered higher compared with other populations (F_{IS}=0.109) like the results reported by Shokrollahi et al., (2009) who found a value of 0.047 in Iranian river buffalo. It could be infer that the obtained low values of genetic variability are consequence of the use of single 70 samples of DNA for the analysis. In our opinion, this it is not a reason of great weight in the obtained results if it is considered that FAO has settled down that for reliable estimation of allele frequencies, at least 25 animals per breed should be typed, but at least 40 animals should be sampled to allow for possible losses, mistyping, missing values and genetic subdivision within breeds or various

degrees of cross-breeding (FAO, 2011). Acosta et al., (2014) found also a lower F_{IS} value (0.087) when studied 16 bovine microsatellite loci in a Cuban buffalo population as well as Martínez et al., (2015) who found low F_{IS} values (0.018) when studied Cuban Buffalypso/Carabao hybrid breed.

The result of the analysis of a possible drastic reduction of heterozygosity showed that in the Cuban population apparently there is equilibrium between mutation and drift, as is show at the graph obtained from the analysis maintains the L-shape (Figure 1 supplementary file). Santana et al., (2011) established that problems exist in the structure of the Murrah buffalo population in southeastern Brazil, in the form of bottlenecks and small effective size; they also identified that inbreeding generally had a negative effect on milk production and quality traits ant the effect observed may have important economic implications for production systems. Moreover, the higher average relatedness coefficient values suggest that inbreeding will continue to quickly increase if breeders do not rapidly implement appropriate management and breeding strategies. That's why we consider that a mating system designed to avoid an increase of inbreeding should be applied to Cuban buffalo population to maintain genetic diversity.

A structure analysis using a Bayesian approach showed the highest ΔK at K=5. The individuals on the buffalo population was separated between them after the first calculation clusters (K=2). Three groups of animals are distinguished in the population, clustered together at K=3 (Figure 2 supplementary file). These results are expected due to the selection applied for genetic improvement of economic traits, mainly milk production and are in agreement with the results obtained by Attia et al., (2014) who investigated biodiversity in Egyptian and Mediterranean buffalo respectively, using microsatellite markers and found significant deviation from HWE. In Cuban population have been no recent imports of live animals despite it has begun to be inseminated with semen from the Mediterranean race to increase genetic diversity introducing new genes to provide better production characteristics. It is confirm the genetic origin of the Cuban population, in which have intervened genes from river breed "Buffalypso" and swamp buffalo (Carabao), plus the recent introduction of genes Mediterranean in the Cuban population.

Results of this study confirm that a large fraction of bovine DNA microsatellite markers can be amplified and is polymorphic in the Cuban buffalo, due to cattle and buffalo species are in a close evolutionary relationship. Also, these DNA markers are applicable for population genetic studies on the Cuban buffalo, that possessed a considerable amount of genetic diversity due to low pressure of artificial selection and possibility of random mating. The Cuban buffalo population require a scientific production system in order to improve the production without losing the significant genetic structure of these economically important animals, so a mating system designed (including an artificial insemination program using out-of-herd animals) to avoid an increase of inbreeding should be applied to Cuban buffalo population to maintain genetic diversity. For further analysis it should be employ the buffalo microsatellite set to study the Cuban population in order to increase the exactitude and specificity on the results. Nevertheless, the present study allowed us to validate the available cattle microsatellite panel in our lab for it use in biodiversity studies in populations of buffalo deepening in the characterization of Cuban one. This tool should be included in the selection strategies of the program of improvement of this species, considering the inbreeding studies to analyze the genetic variability in other way in the Cuban buffalo population.

ACKNOWLEDGEMENTS

This work was supported by research grant MAEC-AECID. We would like to acknowledge Dr. José Raul López Álvarez for providing the blood samples from Buffalo farm of Instituto de Ciencia Animal (ICA), from Cuba and Mrs. Carmen Cons (LAGENBIO) for technical assistance.

REFERENCES

1. **Abou-Bakr S, Ibrahim MAM, Hafez YM, Attia M, Abdel-Salam SM, Mekkawy W** 2012 Genetic characteristics of Egyptian buffalo using DNA microsatellite markers. *Egyptian Journal of Animal Production* **49** 121-126
2. **Acosta A, Uffo O, Sanz A, Obregón D, Ronda R, Osta R, Martin-Burriel I, Rodellar C & Zaragoza P** 2014 Genetic characterization of the Cuban water buffalo population using microsatellite DNA markers. *Buffalo Bulletin* **33** 101-106

3. **Attia M, Abou-Bakr S & Nigm A** 2014 Genetic differentiation and relationship among Egyptian Nile Delta located buffalo using microsatellite markers *Egyptian Journal of Animal Production* **51** 71-77
4. **Borghese A & Mazzi M 2005** Buffalo population and strategies in the world. In: Buffalo Production and Research Rome: FAO; A. Borghese editors. Monterotondo, Roma, Italia
5. **Freeman AR, Bradley DG, Nagda S, Gibson JP, Hanotte O** 2005 Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. Animal Genetics **37** 1-9
6. **Martínez N, da Silva EC, Mitat A, Ponce G, Benício da Silva RC, Paes Barbosa S B, Uffo O & Gomes Filho MA** 2015 Genetic relationships between Cuban and Brazilian buffaloes (*Bubalus bubalis*) by microsatellite markers**.** *Revista Salud Animal* **37** 152-163
7. **Michelizzi VN, Dodson MV, Pan Z, Amaral MEJ, Michal JJ, McLean DJ, Womack JE, & Jiang Z** 2010 Water Buffalo Genome Science Comes of Age. International Journal of Biology Science **6** 333-349
8. **Mitat A** 2009 Búfalos de agua en Cuba. Origen y evolución. Revista ACPA. **3** 45-48
9. **Nagarajan M, Kumar N, Nishanth G, Haribaskar R, Paranthaman K, Gupta J, Mishra M, Vaidhegi R, Kumar S, Rajan AK & Kumar S** 2009. Microsatellites markers of water buffalo, Bubalus bubalis – development, characterization and linkage disequilibrium studies. *BMC Genetics* **10** 68
10. **Santana ML, Aspilcueta-Borquis RR, Bignardi AB, Albuquerque LG & Tonhati H** 2011 Population structure and effects of inbreeding on milk yield and quality of Murrah buffaloes. *J. Dairy Sci.* 94:5204–5211
11. **Shokrollahi B, Amirinia C, Djadid ND, Amirmozaffari N & Ali Kamali M** 2009 Development of polymorphic microsatellite loci for Iranian river buffalo (*Bubalus bubalis*). *African Journal of Biotechnology* **8** 6750-6755
12. **Vieira JN, Teixeira SC, Kuabara MY & Andrade de Oliveira DA** 2011 DNA microsatellites for genetic identification in Brazilian Murrah water buffaloes. *Acta Veterinaria Brasilica* 5 364-367

SUPPLEMENTARY MATERIAL

MATERIALS AND METHODS

Sampling and microsatellite loci, PCR-based profiling, Genetic variability and Genetic Structure are available in a Supplementary File

Peripheral blood samples of 50 adult female buffaloes were collected from a population of unrelated water buffalo (uncontrolled mattings between Buffalypso and Carabao animals), clinically healthy and bred extensively in the Institute of Animal Science (ICA), Mayabeque province, Cuba. Genomic DNA was extracted using Promega Wizard®Genomic DNA purification commercial kit according to the MSRP protocol (Promega Corp., Madison, WI). The quality and quantity of DNA (ng/uL) for each sample was analyzed in a spectrophotometer (Nanodrop ND1000, Thermo Scientific).

A total of 30 heterologous bovine microsatellite loci were chosen for a recommended markers panel the International Society for Animal Genetics (ISAG)/Food and Agriculture Organization of the United Nations (FAO) working group (FAO, 2004).

PCR-based profiling

The polymerase chain reaction (PCR) was carried out using the QIAGEN multiplex PCR kit with 2x QIAGEN multiplex PCR master mix (final concentration, 1x), Q-Solution 5x (final concentration, 0.5x), 0.1 to 0.5 µM of each primer, 20 ng of DNA and distilled water in a total volume of 6 µL. Microsatellite allele sizes were determined with the ABI PRISM 3130 Genetic Analyser (Applied Biosystems, Foster City, CA). The internal size standard GeneScan-500LYS (Applied Biosystems, Warrington, United Kingdom) was used for sizing alleles.

Genetic variability

The GENEPOP package Version 4.0.10 (Raymond and Rousset, 2003) was used to calculate an exact test for deviation from Hardy-Weinberg equilibrium (HWE), allele frequencies, observed and expected heterozygosity. Wright F-statistics (F_{IS}) and overall number of alleles per locus (Na) were calculated using FSTAT (Goudet, 2002). Polymorphism information content was calculated as per Botstein et al., (1980). Inbreeding coefficient in water buffalo population was estimated according to the following equation (Wright, 1965). The program BOTTLENECK (Piry et al., 1999) was used to test whether population analyzed had a reduction in its effective size, developing a temporary heterozygotes excess. Mutation model was applied in two phases (Two-Phases Model, TPM) with variance values for TPM 10, and with a ratio of simple model of mutation (Single Mutation Model, SMM) 90% in this model, with 1000 repetitions, as recommended for most microsatellite loci (Luikart et al., 1998). By assuming this model, deviations from equilibrium drift-mutation were determined following the procedure described by Cornuet and Luikart (1996). The program also it has a quantitative descriptor of the distribution of allele frequencies ("Mode-shift" indicator) which discriminates between stable populations and populations with “bottleneck”.

Genetic structure

The Bayesian model-based method developed by Pritchard et al., (2000) and implemented in the STRUCTURE software was used to investigate population structure and define clusters of individuals on the basis of multi-locus genotypes for 28 microsatellite markers. The number of assumed populations (K) varied between 2 and 10. For each K, 10 independent runs were performed with a burn-in of 10^5 and Monte Carlo Markov Chain (MCMC) length of 10^6 iterations under an admixture and correlated allele frequencies model. The average and standard deviation of the logarithmic likelihood [L(K)] of the data were estimated across 10 runs for each K value. The most probable number of population clusters was determined by plotting L(K) and also using the distribution of ΔK (Evanno et al., 2005). To investigate further population subdivisions, the major clusters identified with STRUCTURE were re-analyzed using the same settings and assuming K=2 to K=n+3 (n being the number of predefined breeds included in each cluster). After assessing the most likely number of underlying populations, the results were graphically displayed with DISTRUCT (available at http://rosenberglab.bioinformatics.med.umich.edu/distruct.html; last accessed June 10, 2016).

Assignment tests were performed with Structure without using prior information of source breeds. The proportion of each individual's genotype in each cluster or breed (*q*) obtained with Structure without using prior information of source breeds were used for assignments. The percentage of individuals correctly assigned to source breeds were calculated for *q*>0.80 and *q*>0.95 thresholds.

Reference of Materials and Methods

3. **Botstein D, White RL, Skolnick M & Davis RW** 1980 Construction of genetic linkage maps in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32** 314- 331
4. **Cornuet JM & Luikart G** 1997 Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144** 2001-2014
5. **Evanno G, Regnaut S & Goudet J** 2005 Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology* **14** 2611-2620.
6. **FAO** 2004 Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Measurement of Domestic Animal Diversity (MoDAD): Recommended Microsatellite Markers. Recommendations of ISAG/FAO Standing Committee, pp. 8–9. http://dad.fao.org (last accessed November 28, 2016)
7. **Goudet J** 2002 FSTAT 2.9.3.2: A program to estimate and test gene diversities and fixation indices, Available at http://www.unil.ch/popgen/softwares/fstat.htm
8. **Luikart G, Allendorf FW, Cornuet JM, Sherwin WB** 1998 Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **9**: 238- 247
9. **Piry S, Luikart G, Cornuet JM** 1999 BOTTLENECK: A program for detecting recent effective population size reductions from allele frequency data. Copyright © 1997-1998, INRA, URLB, Laboratoire de Modélisation et Biologie Evolutive, All rights reserved.
10. **Pritchard JK, Stephens M & Donnelly P** 2000 Inference of population structure using multilocus genotype data. *Genetics* **155** 945-959
11. **Raymond M & Rousset F** 2003 Istruzioni uso GENEPOP e informazioni generali sulle opzioni utilizzabili ("Instruction for GENEPOP use and general information on the usable options"). Available at http://wbiomed.curtin.edu.au/genepop/
12. **Wright S** 1965 The interpretation of population structure by F-statistics with special regard to

systems of mating. *Evolution* **19** 395-420

Supplementary table 1

Table 1. Descriptive statistics of the 27 polymorphic microsatellite marker loci for Cuban water buffalo

Locus	*G*	Na	H_O	H_E	PIC	F_{IS}	*P*-value
BM1818	92	4	0·391	0·339	0·313	−0·156	1·000
TGLA227	100	3	0·640	0·638	0·569	−0·003	0·559
TGLA122	68	6	0·559	0·466	0·458	−0·204	0·965
SPS115	100	7	0·560	0·611	0·531	0·085	0·236
ETH225	94	4	0·298	0·269	0·260	−0·109	1·000
BM2113	100	3	0·440	0·392	0·333	−0·124	0·848
ETH3	92	5	0·413	0·569	0·503	0·277	0·005**
BM1818	98	8	0·400	0·700	0·685	0·431	0·000*
ETH152	100	5	0·580	0·717	0·689	0·192	0·020*
ILSTS006	94	6	0·234	0·661	0·608	0·649	0·000*
INRA23	100	3	0·200	0·184	0·169	−0·090	1·000
CSRM60	98	5	0·612	0·621	0·614	0·014	0·510
INRA05	88	5	0·273	0·444	0·436	0·388	0·001**
INRA63	94	2	0·532	0·395	0·314	−0·353	1·000
HAUT24	74	4	0·108	0·249	0·237	0·569	0·001***
HEL5	90	7	0·267	0·549	0·540	0·517	0·000***
INRA35	98	4	0·327	0·374	0·385	0·127	0·205
HEL9	98	3	0·306	0·357	0·320	0·144	0·173
ILSTS005	98	3	0·469	0·423	0·359	−0·112	0·878
ETH185	98	9	0·551	0·778	0·761	0·292	0·000***
INRA37	92	5	0·348	0·489	0·473	0·292	0·002***
CSSM66	94	8	0·851	0·821	0·809	−0·037	0·728
MM12	92	2	0·109	0·104	0·097	−0·047	1·000
INRA32	46	7	0·435	0·829	0·806	0·481	0·000***
HEL1	92	7	0·805	0·788	0·769	−0·021	0·675
HAUT27	92	7	0·804	0·801	0·784	−0·005	0·606
HEL13	92	8	0·783	0·683	0·674	−0·149	0·965
Mean	91·714	5·036	0·440	0·509	0·482	0·115	0·001**

Na, # alleles; G, # genotype; H_O, observed heterozygosity; HE, expected heterozygosity; PIC, polymorphism information content; F_{IS}, Wright *F*-statistics. Exact test for Hardy–Weinberg equilibrium (*P*-value); statistical significance $*P < 0·05$; $**P < 0·01$; $***P < 0·001$

Supplementary figure S2

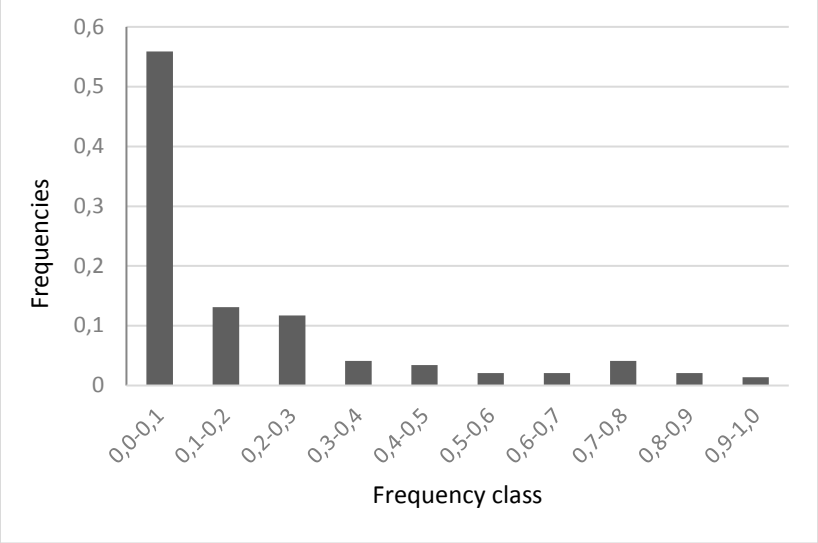

Figure 1. Frequency distribution by classes in the study population