Analysis of microsatellite markers in a Cuban water buffalo breed

Odalys Uffo^{1} et al

SUPPLEMENTARY FILE

SUPPLEMENTARY 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.

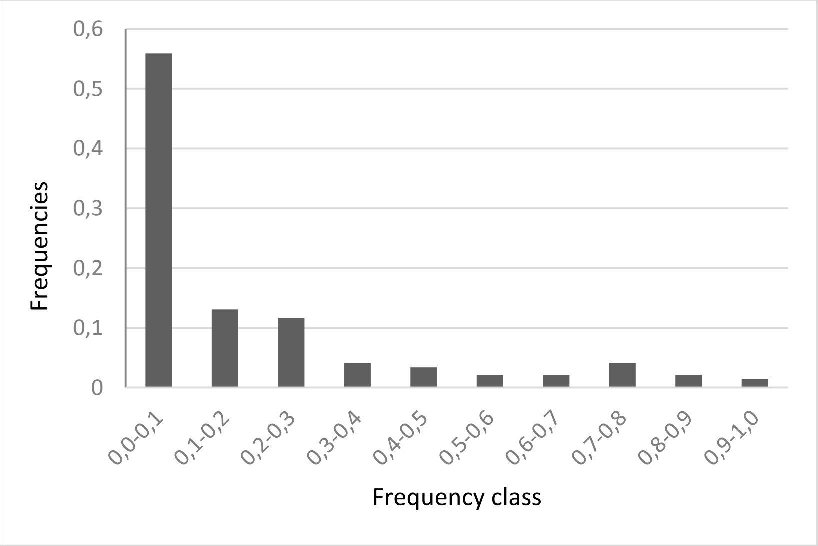
Genetic variability

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.



Supplementary Figure S1. Frequency distribution by classes in the study population

Reference of Materials and Methods

1. **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
2. **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
3. **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.
4. **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)
5. **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
6. **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
7. **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.
8. **Pritchard JK, Stephens M & Donnelly P** 2000 Inference of population structure using multilocus genotype data. *Genetics* **155** 945-959
9. **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/

10. **Wright S** 1965 The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19** 395-420