Animal (2018), **12:3**, pp 464–471 © The Animal Consortium 2017
doi:10.1017/S1751731117001719

animal

Genomic differentiation between swamp and river buffalo using a cattle high-density single nucleotide polymorphisms panel

L. Pérez-Pardal[1a], S. Chen[2a], V. Costa[1], X. Liu[3], J. Carvalheira[1,4] and A. Beja-Pereira[1†]

[1]*Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO-UP), InBIO, Universidade do Porto, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal;* [2]*School of Life Sciences, Yunnan University, No. 2 North Cuihu Road, Kunming 650091, Yunnan Province, China;* [3]*Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming, 650221, Yunnan Province, China;* [4]*Instituto de Ciências Biomédicas de Abel Salazar (ICBAS), University of Porto, Rua de Jorge Viterbo Ferreira 228, 4050-313 Porto, Portugal*

(Received 17 October 2016; Accepted 18 June 2017; First published online 24 July 2017)

*Buffalo (*Bubalus bubalis*) is an important livestock species in many tropical and subtropical regions. In recent decades, the interest in buffalo's milk have expanded and intensive buffalo farms start to emerge. However, breeding programs and population genetics information for this species is scarce or inexistent. The present study aims to test the suitability of the commercial high-density single nucleotide polymorphisms (SNP) genotyping panel, the Illumina BovineHD BeadChip, to estimate population genetics parameters, pedigree control and identification of common variants in major production candidate genes. From a total of 777 962 SNPs included in the panel, 20 479 were polymorphic in water buffalo at a call rate of 86% and an average expected heterozygosity (H_E) of 0.306. From these, 357 were mapped within or around the flanking regions of several major candidate genes. A principal components analysis identified three different clusters, each representing pure swamp buffalo type, pure river buffalo type and admixed river buffalo. The hybrids between swamp and river buffalo were clearly identified as an intermediary cluster. The suitability of these SNPs data set for parentage and identity testing demonstrated that the combination of just 30 to 50 SNPs were enough to attain high probabilities of parentage exclusion (0.9999) in both types and identity (2.3×10^{-5} and 2.0×10^{-7}) for river and swamp buffalo, respectively. Our analysis confirms the suitability of the BovineHD BeadChip to assess population structure, hybridization and identity of the water buffalo populations.*

Keywords: water buffalo, admixture, genome-wide, diversity, traceability

Implications

This study demonstrates the usefulness of a cattle commercial SNPchip panel to estimate several population genetic parameters of in water buffalo. Our results have shown that 20 000 single nucleotide polymorphisms (SNPs) of the nearly 800 000 contained in the SNPchip are polymorphic for the water buffalo. This number of markers was sufficient to clearly distinguish between the swamp and river type buffalos, plus their hybrids. In addition, we have identified a few hundreds of SNPs located in major candidate genes that were associated with milk production traits in cattle. This information can be important to guide selection for milk traits in water buffalo.

Introduction

Water buffalos are an important farm livestock species in many tropical and subtropical climates. Although, traditionally buffalos were raised by small farm holders for multipurpose (milk, draught, meat), recently their use has been expanding to new regions and large and intensive buffalo farms are starting to emerge, particularly for milk production.

Currently, the domestic buffalo is divided into two domestic types with different karyotypes: the swamp type (*Bubalis bubalis carabensis*, $2n = 48$) present in South and Southeast China and Southeast Asia and, the river type (*B. bubalis bubalis*, $2n = 50$) mainly present in India, southwestern Asia, Egypt and southern Mediterranean Europe (Moioli *et al.*, 2001). The origins of the domestic water buffalo types are debated, although phylogenetic studies, mostly based on mitochondrial DNA (mtDNA), indicate that the river type may have originated from India and was domesticated about 5000 years ago in an atypical manner involving continuous introgression of wild animals to the domestic stocks in Indian subcontinent before mature phase of Indus Valley civilization (2600 to 1900 BC) (Nagarajan *et al.*, 2015).

Swamp type may have originated in the border between Indochina and Southwestern China and was domesticated

[a] These authors contribute equally to this work.
[†] E-mail: albanobp@fc.up.pt

https://doi.org/10.1017/S1751731117001719 Published online by Cambridge University Press

about 4000 years ago (e.g. Yang *et al*., 2008; Zhang *et al*., 2016). Y-chromosomal variation also confirms an independent domestication of swamp and river buffalo (Yindee *et al*., 2010).

Swamp buffalo presents a higher diversity of both maternal and paternal lineages than river buffalo and also a strong geographic differentiation. River buffalo has a weak phylogeographic structure (Zhang *et al*., 2016). Hybridization between swamp and river buffalo was also reported (Zhang *et al*., 2016). Crosses between the two types are fertile, although the hybrid may have a lower reproductive value (Harisah *et al*., 1989; Iannuzzi, 2009).

The growing importance of the water buffalo, genome-wide information and the modern sequencing and genotype platforms have been developed (i.e. Axiom® Buffalo Genotyping Array 90 K). Up to now, the number of studies reporting the use of the latter buffalo SNPchip only produced an handful of papers, which were focused at prospecting for major genes associated with milk traits (de Camargo *et al*., 2015; El-Halawanya *et al*., 2017), animal management and breeding (Iamartino *et al*., 2013), population diversity (Colli *et al*., 2016) and a data analysis application (Nicolazzi *et al*., 2014). Two of the major reasons for the small number of papers reporting the use of this technology might be the relatively elevated costs of using this technology and, the existence of a very few breeding programs for this species.

However, water buffalo (*B. bubalis*) and cattle (*Bos taurus*) are phylogenetically very close, have an high chromosomal homology between species and share gene order and chromosome banding homology (Borquis *et al*., 2014). Therefore, buffalo studies can benefit of genomic tools developed for cattle, to gather genome-wide diversity information of the buffalo populations. Thus, the illumina BovineHD BeadChip developed for cattle are yet an interesting tool to collect population genomic information such as variation, structure, origins and relationships between populations and types of buffalos (swamp and river).

As the SNPs incorporated in this array have a uniform coverage of the bovine genome, it is expected to detect major changes induced by selection at coding regions of major candidate genes for milk and meat production, and thus, be as well an important source of genomic information for buffalo breeding programs. As any study using cross-species genetic markers, requires a preliminary testing of its suitability, here we aim study the usefulness of the cattle illumina BovineHD BeadChip to assess population genetics parameters, such as diversity, population structure, levels of population admixture and probabilities of identity (traceability) and parental exclusion.

Material and methods

Sampling and DNA extraction

A total of 15 samples of swamp buffalo and 15 samples of river buffalo were collected in China (ear skin biopsy, Supplementary Material Table S1). Samples from Swamp buffalo lineage selected for meat production were collected in Mangshi, Dehong, Yunnan, while the dairy river buffalo lineage samples were collected in Tengchong, Baoshan, Yunnan. Genomic DNA was extracted using a commercial JETQUICK Tissue DNA Spin Kit extraction (Genomed GmbH, Loehne, Germany) following the standard manufactures' procedure.

As the mtDNA phylogenetic studies revealed the existence of two main distinct maternal lineages (A and B) in swamp buffalo, and one maternal lineage in river buffalo (Lei *et al*., 2007), a 320 base pairs (bp) fragment of mtDNA D-loop were amplified to obtain a preliminary differentiation between river and swamp types, using previously described primers (Troy *et al*., 2001) and procedures for cattle (Beja-Pereira *et al*., 2006).

Samples were genotyped with the Illumina BovineHD BeadChip (Illumina, Inc. San Diego, CA, USA). The array comprises SNPs covering the 29 bovine autosomes, the sex chromosome (BTA X) and three unassigned linkage groups (Matukumalli *et al*., 2009). The BovineHD BeadChip has 777 962 SNP markers spread through the genome, and the average distance between markers of 3.43 kb. Quality control filters were applied in order to eliminate all the genotypes that violated the Hardy–Weinberg equilibrium, to eliminate technology errors of the 'no call' type and monomorphic markers. Filtering was carried out by PLINK (Purcell *et al*., 2007). The QC criteria were minor allele frequency (MAF) threshold of 0.05%, the Hardy–Weinberg equilibrium threshold of $P < 0.001$ and a SNP call rate threshold of 95%. All SNPs were annotated using SnpEff version 3.6 (Cingolani *et al*., 2012) to classify variants (into synonymous and non-synonymous) using the *B. taurus* genome assembly UMD 3.1. Enriched functional annotation clusters were defined using functional annotation tool implemented in Panther (Protein Analysis Through Evolutionary Relationships; www.pantherdb.org/tools/csnp) (Thomas *et al*., 2003).

Statistical analysis

To identify genetic clustering in the buffalo data set we performed principal component analysis (PCA). Vectors constructed with the number of structural variations per chromosome were used to perform a PCA and look for differentiation between buffalo types. We used PLINK command make-rel option square to perform this analysis (Purcell *et al*., 2007). Population assignment analysis was performed using ADMIXTURE (Alexander *et al*., 2009). The program implements a maximum likelihood method to infer the genetic ancestry of each individual from a mixture of *K* predefined ancestral groups. The number of clusters (*K*) tested ranged from 2 to 4. A preferable value of *K* will exhibit a low cross validation error compared with other *K* values.

Admixture between populations was calculated applying the *three-population test* (f3-test) with ADMIXTOOLS (Patterson *et al*., 2012). This is a formal test of admixture, to detect the occurrence of past gene flow events that led into admixture. In brief, what this test does is to test if swamp buffalo type has ancestry from the other two river buffalo lineages and vice versa. If one of the types is unadmixed, then the f3 has non-negative mean. But if f3 has negative

https://doi.org/10.1017/S1751731117001719 Published online by Cambridge University Press

mean this implies that the tested buffalo population is admixed with the other populations (Patterson *et al*., 2012). In our analysis, we divide river buffalo in the two groups identified by the PCA, and by the mtDNA lineage classification (Supplementary Material Table S1). Finally, we also ran these three different tests – Admixture, PCA and f3-test – with a data set containing only the SNPs that were polymorphic in both subspecies and the similarity between these values indicates the consistency of the results (data not shown).

Lastly, since SNPs genotyping profiles have potential to be used in forensic tests of identity and parentage exclusion (for review see e.g. Jobling and Gill, 2004), it can be of extreme importance to correct possible pedigrees errors of those individuals used in breeding programs or property disputes. For this reason, we have estimated the potential power of this set of SNPs to assess the probabilities of identity (*PI*; as the average probability of two independent samples having the same identical genotype) and the probabilities of parentage exclusion (*PE*). This probability was estimated under three different scenarios (Jamieson and Taylor, 1997) for each buffalo type were calculated using GENALEX v.6.5 (Peakall and Smouse, 2012). The probability of exclusion for one putative parent when the other parent's genotype is known $(PE_{X1} = \sum_i^n p_i(1-p)^2 - \sum_{i>j=1}^n (P_iP_j)^2\left[4-3(P_i+P_j)\right])$, where p_i is the frequency of all the alleles $x_{i+\ldots+x_n}$; estimates the probability of parental exclusion when the genotype of one parent is known but the other genotype is unavailable $PE_{X2} = 1-4\sum p_i^2+2\left(\sum p_i^2\right)^2+4\sum p_i^3-3\sum p_i^4$; and the probability of excluding a putative parent pair $(P_{X3} = 1+4\sum p_i^4-4\sum p_i^5-3\sum p_i^6-8\left(\sum p_i^2\right)^2+8\left(\sum p_i^2\right)\left(\sum p_i^3\right)+2\left(\sum p_i^3\right)^2)$. The probability of identity (*PI*) was calculated using a conservative approach that considers the possibility of related individuals be included in the sample, by applying the formula: $PI = 0.25+(0.5\sum p_i^2)+[0.5(\sum p_i^2)^2]-(0.25\sum p_i^4)$ for each locus, where p_i is the allele frequency of a given allele *i* (Waits *et al*., 2001).

Results

A total of 15 samples of swamp buffalo and 15 samples of river buffalo were genotyped with the Illumina BovineHD BeadChip (777 962 SNPs). The genotyping rate was around the 86%. From these SNPs, 145 880 variants were removed due to missing genotype data and 611 603 markers were removed due to minor allele threshold. A total of 20 479 variants (mean gap size = 1269 kb) remained for analysis (Supplementary Material Table S2). For the selected SNPs the genotyping rate was 98% and 724 of this SNPs were not mapped. Mean change rate was one change every 122 666 bases. The observed and expected heterozygosity (H) obtained from this subset of SNPs was $H_O = 0.330$ and $H_E = 0.279$ for the river type (standard deviation (SD) of 0.285 and 0.116, respectively), while for the swamp type it was slightly lower ($H_O = 0.275$ and $H_E = 0.227$, SD of 0.304 and 0.194 correspondingly), for all individuals we found and H_E of 0.306 and $H_O = 0.309$ with a SD of 0.250 and 0.144, respectively.

In river buffalo, 17 572 SNPs were polymorphic and 14 181 SNPs in swamp, and 11 274 SNPs were polymorphic in both subspecies. A total of 6298 SNPs were monomorphic in swamp and 2907 SNPs were monomorphic in river buffalo. All of these SNPs are evenly distributed across the bovine genome (Figure 1). To characterize selected SNPs, we determined the distribution of MAFs in each of the two-domestic subspecies (see Figure 2). The distribution of less frequent allele was slightly different between the two types.

Regarding the final SNPs data set mapping, a total of 357 markers mapped within or around several major candidate genes (Table 1). We have found that some of those SNPs are located in genes related with production traits as growth hormone receptor, prolactin releasing hormone, α-lactalbumin, casein α S1, ATP binding cassette subfamily G member 2 (junior blood group) and diacylglycerol acyltransferases 2; immunological response: BoLA-DRB3 (major histocompatibility complex, class II); and reproduction traits: pregnancy-associated plasma protein A, pappalysin 1 (Santiago *et al*., 2005).

The tests on the possible impact of the SNP changes in the amino acids (aa) and protein function was performed using SnpEff (Cingolani *et al*., 2012), pointed one SNP having high impact, 26 moderate and 265 with low impact, and a total of 22 648 SNPs that despite modifying the aa do not interfere in the protein function (Supplementary Material Table S2). With respect to the protein change, we found 207 synonymous and 26 non-synonymous changes (Supplementary Material Table S2). A total of 11 SNPs affected UTR 5' regions and 57 affected UTR3' regions.

The results of the admixture test revealed two divergent clusters. Admixture between populations is a fundamental process that shapes genetic variation (Patterson *et al*., 2012). At first clustering ($K = 2$), the major ancestral component of swamp is clearly distinct from the major component of river (Figure 3). Only five of 15 river samples do not contain swamp admixture. One swamp individual presented substantial river admixture. At $K = 3$ a separation between river populations was observed. The lowest cross validation error indicates that 2 is the most parsimonious number of clusters.

An exploratory PCA (Figure 4) also revealed two distinct groups. The first principal component (PC1) explained 27.4% of the total variance in the data and separated swamp form river buffalos. Besides of PC1, the only other component explaining a substantial part of the variance was PC2 (≈5%), which separates all the river buffalos in two groups. Regarding the classification based on the mtDNA sequencing data analysis (Supplementary Material Table S1), all swamp buffalos presented lineage A except one individual who presented lineage B. Six river buffalos were from lineage swamp A, one was from lineage swamp B and the rest were identified by the admixture test as from the river lineage.

Finally, the f3-test results obtained after running the analysis on all possible triplet populations combination (river 1, river 2 and swamp; see Supplementary Material Table S1), only the river 2 mtDNA lineage displayed values of admixture ($f3 = -0.0462$ with a Z-score of -19.419). These values,

https://doi.org/10.1017/S1751731117001719 Published online by Cambridge University Press

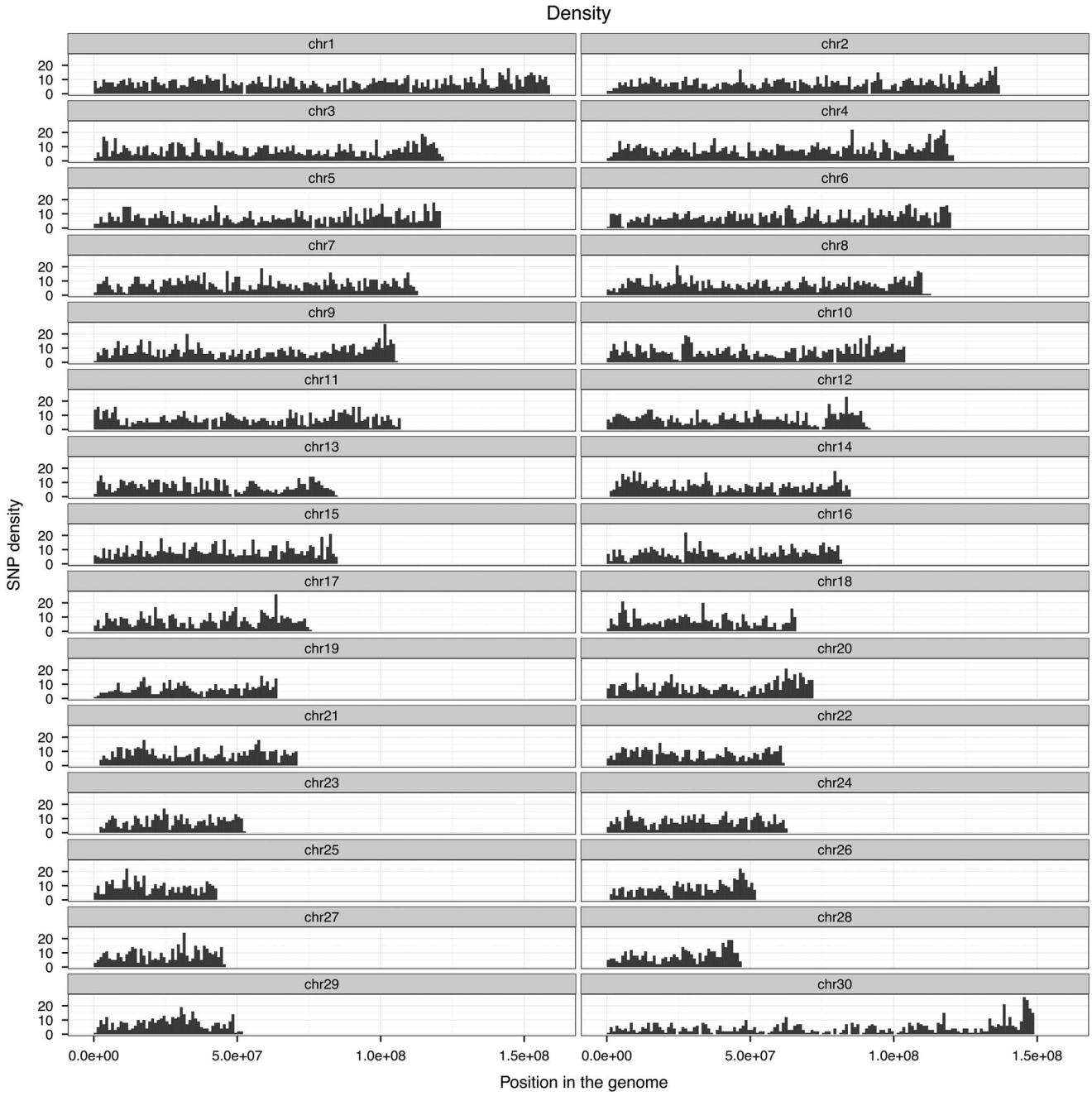


Figure 1 Histogram representing the distribution of the single nucleotide polymorphisms (SNPs) included in our data set (20 479 SNPs) across the bovine genome. Each bar corresponds to 1 million base pairs of the chromosome.

can be taken as a strong evidence that river 2 is composed of both swamp and river type.

The maximum probability of exclusion for a set of 4664 SNPs setting the MAF at ⩾0.3 (Heaton *et al.*, 2014), for the P_{X1} scenario was attained with just 34 and 46 SNPs for river and swamp type, respectively. The maximum probability for the P_{X2} was attained with a combination of 79 and 67 SNPs for river and swamp type, respectively. The maximum probability for the third scenario (P_{X3}) was attained with a combination of 31 and 22 SNPs for river and swamp type, respectively. Finally, the combination of only 40 SNPs attained a *PI* as small as 2.3×10^{-5} and 2.0×10^{-7} for the river and swamp buffalos, respectively.

Discussion

In our study, we found a total of 20 479 SNP polymorphic in water buffalo and a call rate of 86%. In a similar study, 16 580 SNPs were found in 384 swamp type samples at a similar call rate (85%; Borquis *et al.*, 2014). Our higher number of SNPs is mostly due to the use of swamp and river buffalo samples in our testing. In fact, when consider only

https://doi.org/10.1017/S1751731117001719 Published online by Cambridge University Press

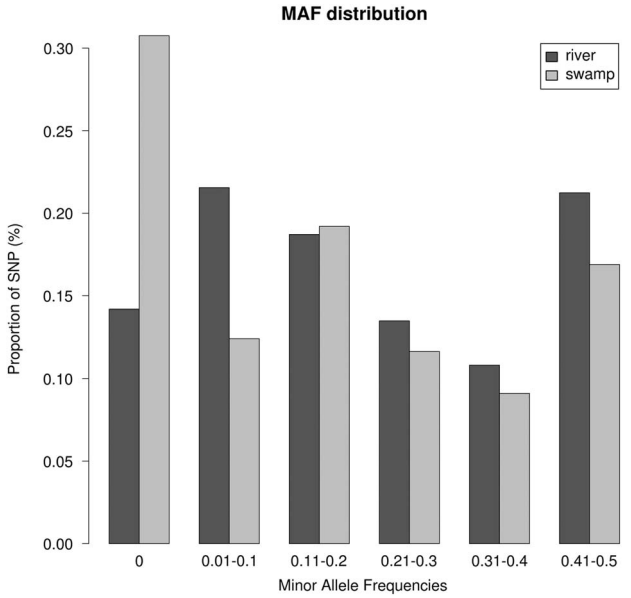


Figure 2 Minor allele frequency (MAF) distributions. Average proportion of single nucleotide polymorphisms (SNPs) of various frequencies for swamp and river buffalo.

the 15 swamp samples, we have found only 14 181 SNPs. When, compared with works where the BovineSNP50 Genotyping BeadChip, were used in water buffalo a much smaller number of sites were found to be polymorphic (around 930) and had a lower call rate (~75%) (Michelizzi *et al*., 2010; Wu *et al*., 2013).

The river buffalo ongoing expansion is generally linked to its dairy production capacity, although they are still used for meat production and draught. Comparative to dairy cattle, dairy buffalo production is still very incipient with a low reproductive efficiency due to long gestation period, long calving interval, silence heat and low conception rate (e.g. see review by Warriach *et al*., 2015). Nonetheless, river buffalo is well suited to poor feeding management as well as adapted to the hot-humid tropical climate condition (Iannuzzi, 2009). As milk production is even lower in swamp buffalo type than in river buffalo, crossings between female swamp buffalos with male river buffalos is increasing with the aim of increasing the milk production in the swamp type (Iannuzzi, 2009). One problem of theses crossings is the lower reproduction value of the hybrids (Harisah *et al*., 1989). Therefore, there a long way to be walked in terms of animal breeding to improve performances at several traits, from which reproduction and milk quantity are the most desired ones. The close genetic proximity with cattle and the allele sharing at major candidate genes, indicate that much of the molecular and physiological architecture of some production traits, might be shared among these two species. Although, in a recent study de Camargo *et al*. (2015) considered that those shared polymorphisms might represent an ancestral polymorphism, and therefore might not be appropriate to study the result of artificial selection in dairy buffalos, we think that SNP frequency distribution patterns in a certain region are more important in

Table 1 *Summarized results of querying the PANTHER database for biological processes, molecular activity and cellular component*

		All SNPs	Non-synonymous
Biological process	Cellular component organization or biogenesis	4.4	2.4
	Cellular process	21.8	22
	Localization	9.9	2.4
	Apoptotic process	1.3	0
	Reproduction	1	2.4
	Biological regulation	11.5	9.8
	Response to stimulus	6.4	7.3
	Developmental process	7.6	7.3
	Rhythmic process	0	0
	Multicellular organismal process	5.6	0
	Locomotion	0.2	0
	Biological adhesion	2.3	9.8
	Metabolic process	24	17.1
	Immune system process	4	7.3
	Cell killing	0	0
Molecular activity	Transporter activity	8.8	5.9
	Translation regulator activity	0.4	0
	Protein-binding transcription factor activity	1.1	0
	Enzyme regulator activity	7.7	5.9
	Catalytic activity	32.2	11.8
	Channel regulator activity	0.1	0
	Receptor activity	10	29.4
	Nucleic acid-binding transcription factor activity	5.8	5.9
	Antioxidant activity	0.1	0
	Structural molecule activity	5.9	11.8
	Binding	27.9	29.4
Celular component	Synapse	0.7	0
	Cell junction	1.1	0
	Membrane	16.6	10
	Macromolecular complex	9.8	0
	Extracellular matrix	3.9	40
	Cell part	37.6	0
	Organelle	22	0
	Extracellular region	8.4	50

PANTHER clustering was performed in two groups: (1) using all data set of SNPs and (2) using only SNPs with non-synonymous effect. Results are given in percentage.

genome-wide association studies than the evolutionary history/genealogy of the SNP itself. For example, a statistical significant allele frequency departure from the neutral expectations, shared by one or more SNPs located in a specific region of the genome might merits to be tested for association with a certain phenotype. Therefore, those shared SNPs which we found located in genes related with production traits, immunological response, and reproduction traits (Supplementary Material Table S2), can be a good starting point for marker assistance selection in buffalo.

Regarding the impact on the biological, cellular and molecular pathways caused by all the analyzed SNPs, the clustering analysis showed that the molecular function most

https://doi.org/10.1017/S1751731117001719 Published online by Cambridge University Press

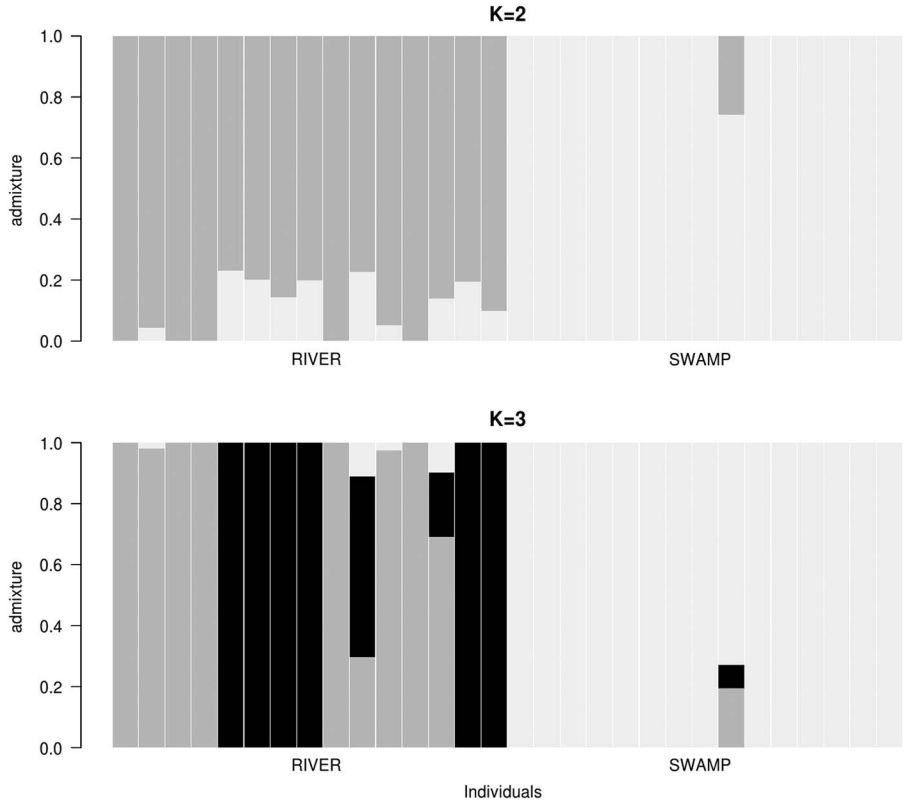


Figure 3 Population structure of swamp and river buffalo. On top admixture plot for $K = 2$, on bottom admixture plot for $K = 3$. Cross validation error (CV error) for: $K = 2$ was 0.47650, $K = 3$ was 0.59353 and for $K = 4$ was 0.65898.

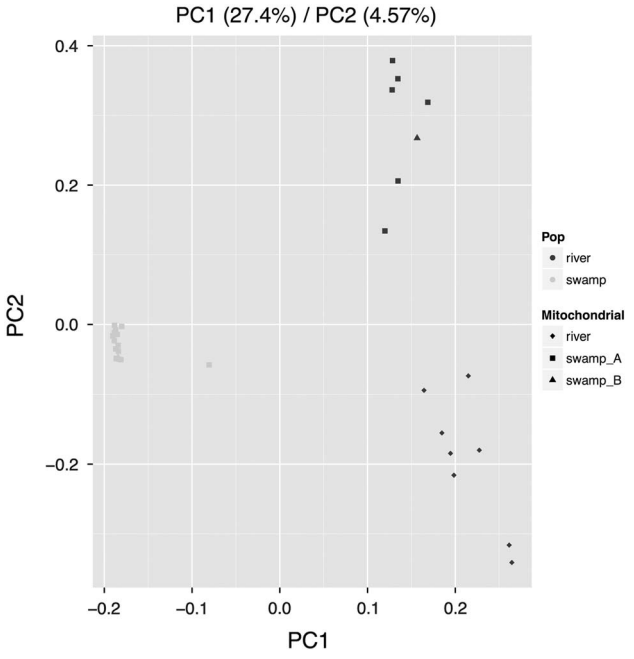


Figure 4 Principal component (PC) analysis. River buffalo samples are in black and swamp in gray. The plot also shows river (diamond), swamp_A (square) and swamp_B (triangle) mitochondrial lineage. All PC values from PC1 to PC10: 27.4, 4.57, 3.48, 3.09, 3.05, 2.97, 2.93, 2.89, 2.84, 2.78.

largely affected are the catalytic and binding functions. When we consider only the non-synonymous aa changes (26 SNPs), the molecular function most largely represented are the binding and the receptor activity (Supplementary Material Table S2). When considering both (all SNPs plus the non-synonymous changes) the most common processes affected were the cellular process and the metabolic process. It is important to note that it is not uncommon for a gene to have more than one transcript and consequently the same variant (e.g. a SNP) might affect different transcripts in different modes. Therefore, careful should be paid when interpreting the different scales of impact estimated by the method used by SNPeff software for a given SNP.

The buffalo classification based on the mtDNA revealed a striking alignment with the results obtained from the admixture and PC analysis. The genetic variation of swamp buffalo is clearly distinct from river buffalo, and both our results, mtDNA sequencing and SNPchip, suggest that the three main groups (swamp, river 1 and river 2, Supplementary Material Table S1) can be differentiated from each other. The negative f3-test results can be taken as a strong evidence that river 2 is composed of both swamp and river types. Indeed, this analysis showed that nearly half of the river individuals were crossed with swamp buffalos (seven out of 15; see Figures 3 and 4).

Finally, the P_E and P_I estimates clearly demonstrate that a combination of a small number of SNPs (<80) offers enough power to be used in parental testing and traceability (identity). The slightly different P_E and P_I values obtained for the different types reflect the heterozygosity differences observed at each buffalo type – river and swamp. Nonetheless, none of the

https://doi.org/10.1017/S1751731117001719 Published online by Cambridge University Press

combinations necessary to obtain the highest probabilities required more than 80 SNPs. Therefore, this set of SNPs can be used for testing parentage and traceability of the many breeds and types of water buffalos.

Conclusions

Here we demonstrated that a subset of 20 000 SNPs from the BovineHD BeadChip is sufficient to differentiate between swamp and river buffalo and to estimate the different admixture proportion in crosses of both types. However, a denser SNP data set would permit to increase the accuracy of the genomic linkage disequilibrium and selective sweeps for these species. Nonetheless, the current number of SNPs (±20 000) found in this SNPchip, has enough information to draw some population genetics parameters (parentage and individual identification), and the fact that it includes SNPs located in major candidate genes involved in milk production traits, it can be useful in further studies to determine whether these SNPs are significantly associated with these traits in buffalos as they are in dairy cattle.

Acknowledgments

This work was supported by a project grant from FCT PTDC/CVT/117851/2010. LP-P is funded by an FCT scholarship (SFRH/BPD/94518/2013). AB-P is supported by FCT through an IF-FCT contract (IF/00370/2013). V.C. is funded by an FCT scholarship (SFRH/BD/88129/2012) and S.C. is supported by High-Level Overseas Talents Recruitment Program of Yunnan Province. The authors thank the two anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731117001719

References

Alexander DH, Novembre J and Lange K 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19, 1655–1664.

Beja-Pereira A, Caramelli D, Lalueza-Fox C, Vernesi C, Ferrand N, Casoli A, Goyache F, Royo LJ, Conti S, Lari M, Martini A, Ouragh L, Magid A, Atash A, Zsolnai A, Boscato P, Triantaphylidis C, Ploumi K, Sineo L, Mallegni F, Taberlet P, Erhardt G, Sampietro L, Bertranpetit J, Barbujani G, Luikart G and Bertorelle G 2006. The origin of European cattle: evidence from modern and ancient DNA. Proceedings of the National Academy of Sciences of the United States of America 103, 8113–8118.

Borquis RR, Baldi F, de Camargo GM, Cardoso DF, Santos DJ, Lugo NH, Sargolzaei M, Schenkel FS, Albuquerque LG and Tonhati H 2014. Water buffalo genome characterization by the Illumina BovineHD BeadChip. Genetics and Molecular Research 13, 4202–4215.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X and Ruden DM 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6, 80–92.

Colli L, Milanesi M, Vajana E, Iamartino D, Bomba L, Puglisi F, Del Corvo M, Nicolazzi E, Ahmed SSE-d, Herrera JRV, Cruz L, Zhang S, Yang BZ, Hao X, Zuo F, Lai S-J, Wang S, Liu R, Gong Y, Mokhber M, Shahrbabak HM, Mao Y, Guan F, Vlaic A, Ramunno L, Ahmad A, Soysal I, Ünal EÖ, Ketudat-Cairns M, Garcia JF, Utsunomiya YT, Parnpai R, Drummond MG, Galbusera P, Burton J, Hoal E, Yusnizar Y, Sumantri C, Moioli B, Valentini A, Stella A and Williams J, International Buffalo Consortium and Ajmone Marsan P 2016. Water buffalo genomic diversity and post-domestication migration routes. In Plant & Animal Genome Conference XXIV, San Diego, USA.

de Camargo GMF, Aspilcueta-Borquis RR, Fortes MRS, Porto-Neto R, Cardoso DF, Santos DJA, Lehnert SA, Reverter A, Moore S and Tonhati H 2015. Prospecting major genes in dairy buffaloes. BMC Genomics 16, 872.

El-Halawanya N, Abdel-Shafy H, Shawkya A-E-MA, Abdel-Latifc MA, Al-Tohamya AFM and El-Moneima OMA 2017. Genome-wide association study for milk production in Egyptian buffalo. Livestock Science 198, 10–16.

Harisah M, Azmi TI, Hilmi M, Vidyadaran MK, Bongso TA, Nava ZM, Momongan V and Basrur PK 1989. Identification of crossbred buffalo genotypes and their chromosome segregation patterns. Genome 32, 999–1002.

Heaton MP, Leymaster KA, Kalbfleisch TS, Kijas JW, Clarke SM, McEwan J, Maddox JF, Basnayake V, Petrik DT, Simpson B, Smith TP and Chitko-McKown CG, International Sheep Genomics C 2014. SNPs for parentage testing and traceability in globally diverse breeds of sheep. PLoS One 9, e94851.

Iamartino D, Williams JL, Sonstegard T, Reecy J, Cv Tassell, Nicolazzi EL, Biffani S, Biscarini F, Schroeder S, de Oliveira DAA, Coletta A, Garcia JF, Ahmad A, Ramunno L, Pasquariello R, Drummond MG, Bastianetto E, Fritz E and Knoltes J 2013. The buffalo genome and the application of genomics in animal management and improvement. In 10th World Buffalo Congress and 7th Asian Buffalo Congress, Phuket, Thailand, pp. 151–158.

Iannuzzi LDM, Di Meo G 2009. Water buffalo. In Genome mapping and genomics in domestic animals. (eds. NE Cockett and C Kole), pp. 19–28, Springer-Verlag, Berlin Herdelberg, Germany.

Jamieson A and Taylor SC 1997. Comparisons of three probability formulae for parentage exclusion. Animal Genetics 28, 397–400.

Jobling MA and Gill P 2004. Encoded evidence: DNA in forensic analysis. Nature Reviews in Genetics 5, 739–751.

Lei CZ, Zhang W, Chen H, Lu F, Ge QL, Liu RY, Dang RH, Yao YY, Yao LB, Lu ZF and Zhao ZL 2007. Two maternal lineages revealed by mitochondrial DNA D-loop sequences in chinese native water buffaloes (*Bubalus bubalis*). Asian-Australasian Journal of Animal Sciences 20, 471–476.

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS and Van Tassell CP 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4, e5350.

Michelizzi VN, Wu X, Dodson MV, Michal JJ, Zambrano-Varon J, McLean DJ and Jiang Z 2010. A global view of 54,001 single nucleotide polymorphisms (SNPs) on the Illumina BovineSNP50 BeadChip and their transferability to water buffalo. International Journal of Biological Sciences 7, 18–27.

Moioli B, Georgoudis A, Napolitano F, Catillo G, Giubilei E, Ligda C and Hassanane M 2001. Genetic diversity between Italian, Greek and Egyptian buffalo populations. Livestock Production Science 70, 203–211.

Nagarajan M, Nimisha K and Kumar S 2015. Mitochondrial DNA variability of domestic river buffalo (*Bubalus bubalis*) populations: genetic evidence for domestication of river buffalo in Indian subcontinent. Genome Biology and Evolution 7, 1252–1259.

Nicolazzi EL, Iamartino D and Williams JL 2014. AffyPipe: an open-source pipeline for affymetrix axiom genotyping workflow. Bioinformatics 30, 3118–3119.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T and Reich D 2012. Ancient admixture in human history. Genetics 192, 1065–1093.

Peakall R and Smouse PE 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28, 2537–2539.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81, 559–575.

Santiago CA, Voge JL, Aad PY, Allen DT, Stein DR, Malayer JR and Spicer LJ 2005. Pregnancy-associated plasma protein-A and insulin-like growth factor binding protein mRNAs in granulosa cells of dominant and subordinate follicles of preovulatory cattle. Domestic Animals Endocrinology 28, 46–63.

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A and Narechania A 2003. PANTHER: a library of

https://doi.org/10.1017/S1751731117001719 Published online by Cambridge University Press

protein families and subfamilies indexed by function. Genome Research 13, 2129–2141.

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes BC and Bradley DG 2001. Genetic evidence for near-eastern origins of European cattle. Nature 410, 1088–1091.

Waits LP, Luikart G and Taberlet P 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology 10, 249–256.

Warriach HM, McGill DM, Bush RD, Wynn PC and Chohan KR 2015. A review of recent developments in buffalo reproduction – a review. Asian Australasian Journal of Animal Science 28, 451–455.

Wu JJ, Song LJ, Wu FJ, Liang XW, Yang BZ, Wathes DC, Pollott GE, Cheng Z, Shi de S, Liu QY, Yang LG and Zhang SJ 2013. Investigation of transferability of BovineSNP50 BeadChip from cattle to water buffalo for genome wide association study. Molecular Biology Reports 40, 743–750.

Yang DY, Liu L, Chen X and Speller CF 2008. Wild or domesticated: DNA analysis of ancient water buffalo remains from north China. Journal of Archaeological Science 35, 2778–2785.

Yindee M, Vlamings BH, Wajjwalku W, Techakumphu M, Lohachit C, Sirivaidyapong S, Thitaram C, Amarasinghe AA, Alexander PA, Colenbrander B and Lenstra JA 2010. Y-chromosomal variation confirms independent domestications of swamp and river buffalo. Animal Genetics 41, 433–435.

Zhang Y, Lu Y, Yindee M, Li KY, Kuo HY, Ju YT, Ye S, Faruque MO, Li Q, Wang Y, Cuong VC, Pham LD, Bouahom B, Yang B, Liang X, Cai Z, Vankan D, Manatchaiworakul W, Kowlim N, Duangchantrasiri S, Wajjwalku W, Colenbrander B, Zhang Y, Beerli P, Lenstra JA and Barker JS 2016. Strong and stable geographic differentiation of swamp buffalo maternal and paternal lineages indicates domestication in the China/Indochina border region. Molecular Ecology 25, 1530–1550.

https://doi.org/10.1017/S1751731117001719 Published online by Cambridge University Press