R. Bras. Zootec., 49:e20190267, 2020
https://doi.org/10.37496/rbz4920190267

Breeding and genetics
Full-length research article

Brazilian Journal of Animal Science
e-ISSN 1806-9290
www.rbz.org.br

Identification of candidate genes associated with milk yield trait in buffaloes (*Bubalus bubalis*) by restriction-site-associated DNA sequencing

Manhong Ye[1,2], Mengting Xu[1,3], Manran Lu[1,3], Bin Zhou[4*], Heba Abd El-Kader[5], Sally Said Alam[5], Karima Fathy Mahrous[5]

[1] College of Bioscience and Biotechnology, Yangzhou University, Yangzhou, China.
[2] Key Laboratory of Animal (Poultry) Genetics Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, Beijing, China.
[3] Joint International Research Laboratory of Agricultural & Agri-Product Safety, the Ministry of Education of China, Yangzhou University, Yangzhou, Jiangsu, China.
[4] College of Animal Science and Technology, Yangzhou University, Yangzhou, China.
[5] Department of Cell Biology, Genetic Engineering and Biotechnology Division, National Research Center, Dokki, Giza, Egypt.

***Corresponding author:**
bzhou@yzu.edu.cn

Received: December 13, 2019
Accepted: May 26, 2020

How to cite: Ye, M.; Xu, M.; Lu, M.; Zhou, B.; Abd El-Kader, H.; Alam, S. S. and Mahrous, K. F. 2020. Identification of candidate genes associated with milk yield trait in buffaloes (*Bubalus bubalis*) by restriction-site-associated DNA sequencing. Revista Brasileira de Zootecnia 49:e20190267.
https://doi.org/10.37496/rbz4920190267

Copyright: This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT - The objectives of our present study included the screening of single nucleotide polymorphisms (SNP) that show significant differences in allelic frequencies between two buffalo populations (Egyptian and Chinese buffaloes), categorization of functional genes associated with these SNP by gene ontology, and pathway analyses to further understand their potential values as candidate genes closely associated with milk yield trait in buffaloes. In this study, double digest restriction-site associated DNA sequencing was performed on Illumina HiSeq 2500 platform for 20 and 25 female buffaloes from Egypt and China, respectively. Approximately 118 Gb of sequencing data were obtained, and a total of 110,129 and 150,535 putative SNP were detected in Egyptian and Chinese buffaloes, respectively. Focused only on those SNP that differed significantly in allelic frequencies between the two populations, we found that genes associated with these SNP were significantly over-represented in the ionotropic glutamate receptor pathway, the endothelin signaling pathway, and the gonadotropin-releasing hormone receptor pathway, which contained a total of 29 genes. Of these, nine genes (*ADCY5*, *CACNA1A*, *CREB1*, *INHBA*, *INHBB*, *PIK3R1*, *PLCB1*, *PRKCE*, and *SMAD2*) participating in the hormonal regulation of lactation, were considered to be promising candidate genes worthy of further investigations for favorable alleles associated with milk yield. Our results provide useful information about genetic variations in Egyptian and Chinese buffaloes. The potential influences of nine candidate genes and their associated SNP on milk yield need to be validated in more buffalo populations.

Keywords: association, functional gene, pathway analysis, SNP screening

1. Introduction

Water buffalo (*Bubalus bubalis*) is an important source of milk in Egypt and China. Buffalo milk is well known for its high milk qualities. Compared with cow milk, it has higher contents of fat, protein, lactose, and minerals and is more suitable for the manufacture of various dairy products (Michelizzi et al., 2010). The Egyptian buffalo (EGB) and the indigenous Chinese buffalo (CHB) belong to two different buffalo subspecies, the river buffalo and the swamp buffalo, respectively. They differ significantly in milk yield, approximately 2200-2400 kg/year for the riverine buffalo, and 500-700 kg/year for the swamp buffalo (Shi et al., 2012). Along with the increasing demand for

high-quality buffalo milk, increasingly more buffalo-breeding programs are aiming for buffaloes with improved milk performance, especially those with increased milk yield.

Marker-assisted selection is a useful approach to assist animal breeding. Recently, candidate genes harboring single nucleotide polymorphisms (SNP) significantly associated with milk fat content (Li et al., 2018c; Gu et al., 2019), protein percentage (Manzoor et al., 2018, 2020), and fatty acid composition (Cosenza et al., 2017, 2018) have been identified on different buffalo chromosomes. As for milk yield, one of the most important economic traits in buffalo industry, associated polymorphisms have been identified at the whole genome level (Wu et al., 2013; Venturini et al., 2014; de Camargo et al., 2015; Menon et al., 2016; El-Halawany et al., 2017; Iamartino et al., 2017; da Costa Barros et al., 2018; Du et al., 2019). These studies have made great contributions to the progress of various breeding programs in buffalo. The continuous development of molecular technology enables researchers to comprehensively study the genome of buffalo and, thus, provides more useful information for buffalo breeding.

With the development of next-generation sequencing techniques, information for the whole buffalo genome sequence is now available at https://www.ncbi.nlm.nih.gov/assembly/GCA_003121395.1. Restriction site-associated DNA (RAD) sequencing (RAD-Seq) is one of the next-generation sequencing techniques that have been deployed for detection of large number of SNP quickly and inexpensively (Peterson et al., 2012). To our best knowledge, the RAD approach has not been applied in buffaloes to identify sequence polymorphisms.

In this study, we performed double digest (dd) RAD-Seq to detect genome-wide SNP in EGB and CHB populations with extreme difference in milk yield. Unlike the Egyptian buffalo, which has a relatively high milk yield, the Chinese buffalo, Haizi buffalo, is traditionally used for labor in farmlands with very low milk yield that is barely enough for the consumption of the calf. It is one of the most famous local buffalo breeds in China for its strong adaptability to hard environments and tasty meat. However, with the development of agriculture mechanization, Haizi buffalo is no longer needed in agricultural farming and was at the edge of extinction with only 1,132 head of live buffaloes in 2006 (Cheng et al., 2008). The marker-assisted selection programs, aiming at improving the milk yield trait in Haizi buffalo and thus facilitating its conversion from draft-purpose only to meat-and-milk use, provide a practical way for the conservation of this unique buffalo breed.

The objectives of our present study included the screening of SNP that show significant differences in allelic frequencies between these two populations, categorization of functional genes associated with these SNP by gene ontology (GO), and pathway analyses to further understand their potential values as candidate genes closely associated with milk yield trait in buffaloes.

2. Material and Methods

Blood samples from 45 two-year-old female buffaloes (25 and 20 individuals from the Chinese Haizi swamp buffalo and Egyptian river buffalo, respectively) were collected. The geographic sites were Yancheng, Jiangsu Province, China (33°50' N, 120°22' E) and Cairo, Egypt (30°03' N, 31°58' E), respectively.

Genomic DNA (gDNA) was extracted from the whole blood using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. The quality of gDNA was verified on 1% agarose gel and then quantified using a spectrophotometer (Nanodrop 2000, Thermo Fisher Scientific, Waltham, MA, USA).

The ddRAD-Seq library was constructed by staff in Personal Biotechnology Co., Ltd (Shanghai, China) according to Peterson et al. (2012). Enzymes used in this study were purchased from the New England Biolabs (NEB), Beverly, MA, USA. Briefly, 500 ng of gDNA was double-digested with *Hind*III (5'- A|AGCTT -3') and *Bfa*I (5'- C|TAG -3') at 37 °C for 3 h in a reaction volume of 20 µL. The digested gDNA was purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). Then, in a 50-µL ligation system, adapter P1 (containing a unique 7-bp barcode sequence and a *Hind*III restriction site overhang) and adapter P2 (containing a *Bfa*I overhang) were ligated to the digested gDNA by using T4 DNA ligase. Ligation was performed at 22 °C for 60 min, followed by an inactivation step of 65 °C

for 10 min. The resulting samples were purified by following the standard AMPure XP bead protocol to remove unligated adapters and adapter-adapter ligation products.

After the purification step, the fragments were screened using the Pippin Prep system (Sage Science, Beverly, MA, USA) to obtain 200-400 bp fragments. Enrichment of PCR of the library was performed using a Phusion High-Fidelity PCR Kit (NEB, Beverly, MA, USA) in a 20-μL PCR reaction volume containing about 20 ng of the size-selected DNA as the template and 2 μM each of P1 and P2 adapter primers. The PCR conditions were: initial denaturation at 98 °C for 30 s, then 14 cycles of 98 °C for 15 s, 65 °C for 30 s, and 72 °C for 30 s, followed by a final extension step at 72 °C for 5 min. The PCR samples were purified again with AMPure XP beads and checked on an Agilent Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA, USA) to confirm the size distribution of fragments, and then quantified using a Qubit 3.0 fluorometer (Life Invitrogen, USA). Samples were then combined in equimolar ratios and sequenced on the platform of Illumina HiSeq 2500.

According to the 7-bp barcode sequences assigned to each sample in the ddRAD-Seq library, raw Illumina reads were de-multiplexed, which allowed the matching of each sequence read to a single sample. High-quality (HQ) cleansed reads were obtained by following these filtering criteria: trimming out the barcode sequences as well as the adapter sequences; removing reads with more than 50% bases having Phred quality scores lower than 20; and eliminating the pair-end sequences shorter than 50 bp. Then, the HQ reads were mapped to the reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCA_003121395.1).

The Burrows-Wheeler Aligner (BWA) tool (v0.7.12) was used to align cleansed reads against the reference buffalo genome (GCA_003121395.1 UOA_WB_1) with the demo BWA mem settings. Variant calling was performed according to Zhu et al. (2018). The SNP were further screened by using vcftools based on these criteria: minor allele frequency (considering 45 samples together) ≥0.01; minimum depth of coverage for each sample ≥2; SNP missing rate across all samples (computed per population) ≤0.7; minimum Phred quality score for each bases ≥10; and max alleles/min alleles ≤2.

The SNP, successfully genotyped in all the 45 samples (missing genotype rate = 0) with a minimum allele frequency of 0.01 in both CHB and EGB, were retained for further bioinformatics analysis. Differences in allelic frequencies between CHB and EGB were calculated for all the polymorphic loci using the Population Differentiation option of GenePop version 4.2 (http://genepop.curtin.edu.au/). Significant differences were established at P<0.05. Gene ontology and pathway analyses were performed by the PANTHER classification system (v.14.0) (Mi et al., 2019). We used the "Gene List Analysis" tool. Briefly, the list of genes (official gene symbols) was uploaded, and the *Bos taurus* reference was selected as the reference list. For four functional categories (GO biological process complete, GO molecular function complete, GO cellular component complete, and PANTHER pathways), Fisher's Exact test (calculating False Discovery Rate, FDR) was applied to determine whether there was a statistical over- or under-representation of genes/proteins in the input list relative to the reference list (Mi and Thomas, 2009). The interactions between proteins encoded by the promising candidate genes were predicted using the software Strings (http://string-db.org/).

Raw sequence data obtained from the present study were deposited in the NCBI Sequence Read Archive under the BioProject number of PRJNA554744 (SRA accession: SRR9831104- SRR9831148).

3. Results

In the present study, the Illumina HiSeq sequencing of the 45 ddRAD libraries generated a total of 833.68 million raw reads corresponding to about 118 Gb of sequence data. The average number of raw reads per sample was 18,526,242, ranging from 13,094,116 to 25,602,402 with a median of 19,117,758 and standard deviation (SD) of 3,029,678. After trimming and filtering, we obtained approximately 797.23 million (95.63%) HQ reads, of which, 99.24 and 97.10% of the bases had a Phred quality score of 20 (Q20) and 30 (Q30), respectively. The average number of HQ reads per sample was 17,716,223, ranging from 12,532,968 to 23,686,928 with a median of 17,850,416 and SD of 2,984,099. High-quality reads were then aligned with the reference buffalo genome, and about 99.93% of them were successfully mapped. The average sequencing depth was 0.86X. Summary of the ddRAD-Seq results were shown in Table 1.

Table 1 - Summary of double digest restriction site-associated DNA (ddRAD) sequencing

Sample ID		Raw data					High-quality data					Mapping%	Sequencing depth
		Read number	Total bases	GC%	Q20	Q30	Read number	Total bases	GC%	Q20	Q30		
Chinese swamp buffalo (CHB)	CHB1	23483262	3345991045	51.55	98.22	95.24	22666712	3178681630	51.50	99.36	97.34	99.94	1.11
	CHB10	13660992	1946820967	51.21	98.38	95.55	13212758	1856249624	51.16	99.39	97.45	99.93	0.63
	CHB11	22154648	3156451570	51.31	98.35	95.49	21428862	3008829863	51.26	99.38	97.42	99.94	1.03
	CHB12	20794648	2962766393	50.63	97.03	92.56	19269706	2673377743	50.56	98.85	95.94	99.93	0.97
	CHB13	17116116	2438646583	50.49	97.12	92.73	15898504	2207722249	50.42	98.86	95.98	99.93	0.79
	CHB14	19141986	2727510678	50.71	97.09	92.65	17749796	2465049223	50.64	98.86	95.97	99.93	0.89
	CHB15	18829828	2683148163	51.22	97.05	92.58	17471544	2424334962	51.15	98.85	95.92	99.93	0.87
	CHB16	19117758	2724260348	50.65	97.09	92.68	17747388	2464321459	50.58	98.86	95.98	99.92	0.89
	CHB17	19352994	2757680662	49.68	96.91	92.31	17793190	2470113665	49.60	98.81	95.84	99.92	0.92
	CHB18	14601084	2080212343	50.43	97.01	92.55	13503522	1874404644	50.36	98.85	95.96	99.93	0.69
	CHB19	15824808	2254549289	50.90	97.10	92.72	14688264	2039401837	50.83	98.88	96.04	99.92	0.75
	CHB2	20021100	2852797991	51.52	98.25	95.30	19330432	2711810289	51.46	99.37	97.36	99.93	0.95
	CHB20	15554564	2216323698	50.62	97.05	92.59	14414186	2000938996	50.55	98.85	95.95	99.93	0.72
	CHB21	13565544	1932895934	50.64	96.94	92.38	12532968	1737769606	50.56	98.84	95.90	99.92	0.63
	CHB22	16781692	2391529428	50.67	97.01	92.53	15520054	2154991876	50.60	98.86	95.96	99.93	0.78
	CHB23	14087196	2007059441	50.85	97.33	93.21	13163216	1832286401	50.79	98.93	96.21	99.92	0.66
	CHB24	25602402	3647828885	50.84	97.04	92.57	23686928	3288141585	50.78	98.85	95.95	99.92	1.18
	CHB25	20004772	2850526487	50.89	97.11	92.71	18566866	2579046583	50.82	98.87	95.98	99.93	0.93
	CHB3	22209640	3164493855	51.16	98.17	95.11	21404240	3000276126	51.10	99.35	97.28	99.93	1.08
	CHB4	21184982	3018436318	51.06	98.33	95.45	20479252	2875244683	51.01	99.38	97.41	99.94	1.01
	CHB5	21044612	2998680228	51.56	98.19	95.16	20291230	2844179769	51.50	99.35	97.30	99.93	1.00
	CHB6	14901502	2122978529	51.39	98.28	95.36	14403006	2020465701	51.33	99.37	97.39	99.93	0.71
	CHB7	18503152	2636571988	51.25	98.20	95.19	17850416	2502434069	51.20	99.35	97.32	99.93	0.89
	CHB8	17669184	2517594782	51.06	98.26	95.32	17067010	2394258215	51.01	99.36	97.35	99.93	0.86
	CHB9	22367020	3187119649	51.01	98.36	95.52	21650772	3040402821	50.96	99.39	97.42	99.92	1.07

Continues...

Table 1 (Continued)

Sample ID		Raw data					High-quality data					Mapping%	Sequencing depth
		Read number	Total bases	GC%	Q20	Q30	Read number	Total bases	GC%	Q20	Q30		
Egyptian river buffalo (EGB)	EGB1	19954660	2842865676	50.99	98.50	95.83	19364892	2721824674	50.95	99.42	97.57	99.92	0.91
	EGB10	13094116	1865663266	52.41	98.80	96.69	12794148	1802969450	52.38	99.53	98.06	99.92	0.60
	EGB11	17304894	2465864004	52.98	98.72	96.5	16873112	2377092551	52.94	99.51	97.97	99.94	0.79
	EGB12	18741238	2670639359	52.84	98.77	96.61	18296674	2578620130	52.80	99.52	98.02	99.92	0.83
	EGB13	16020408	2282435287	52.48	98.75	96.59	15642202	2203598697	52.45	99.51	98.01	99.94	0.74
	EGB14	21578934	3074784478	55.86	98.66	96.36	21011628	2958560798	55.82	99.49	97.90	99.94	0.99
	EGB15	19704988	2807575460	53.25	98.80	96.68	19249162	2713243995	53.21	99.52	98.03	99.92	0.87
	EGB16	20921572	2981120627	52.75	98.75	96.57	20420138	2877061205	52.72	99.51	97.99	99.95	0.94
	EGB17	19359784	2758187364	51.23	98.66	96.36	18860878	2654363628	51.19	99.49	97.91	99.95	0.92
	EGB18	18513462	2637721803	52.85	98.75	96.57	18063298	2544782377	52.82	99.52	98.02	99.93	0.82
	EGB19	14351804	2044907188	50.79	98.34	95.69	13875576	1948100342	50.76	99.39	97.63	99.94	0.69
	EGB2	14196368	2022727385	51.51	98.28	95.34	13697516	1922456015	51.47	99.33	97.32	99.94	0.65
	EGB20	19359496	2758848086	53.79	98.57	96.18	18818842	2648032249	53.75	99.47	97.84	99.92	0.90
	EGB3	17859996	2544885691	51.09	98.34	95.46	17276294	2425544690	51.03	99.38	97.41	99.93	0.83
	EGB4	13471634	1919720732	53.07	97.30	93.29	12689384	1766386475	53.04	99.16	96.65	99.94	0.63
	EGB5	17726002	2526040780	50.18	97.32	93.35	16692040	2324816545	50.15	99.21	96.77	99.94	0.86
	EGB6	20876454	2974128637	56.28	98.62	96.26	20319050	2858790994	56.24	99.47	97.84	99.89	0.97
	EGB7	20921442	2980137664	56.10	98.68	96.39	20387808	2869809003	56.06	99.49	97.88	99.90	0.95
	EGB8	20532550	2924949190	51.17	98.65	96.36	20009496	2815285014	51.14	99.49	97.92	99.94	0.99
	EGB9	21615612	3079790305	51.77	98.75	96.59	21097104	2971753216	51.74	99.52	98.03	99.95	0.99

A total of 110,129 and 150,535 putative SNP were identified in Egyptian and Chinese buffaloes, respectively. The distribution of these SNP across the chromosomes was similar within the two populations (Figure 1). Most of them were distributed in the intergenic and intronic regions (Figure 2). The flow chart for SNP detection and bioinformatics analysis were explained in Figure 3.

Altogether, 608 synonymous and 541 non-synonymous mutations were detected on the exons of functional genes. Of all these 1149 SNP, we focused only on high-confidence SNP with no missing data among all the 45 samples, having a minimum reads coverage depth of 2 and showing significant

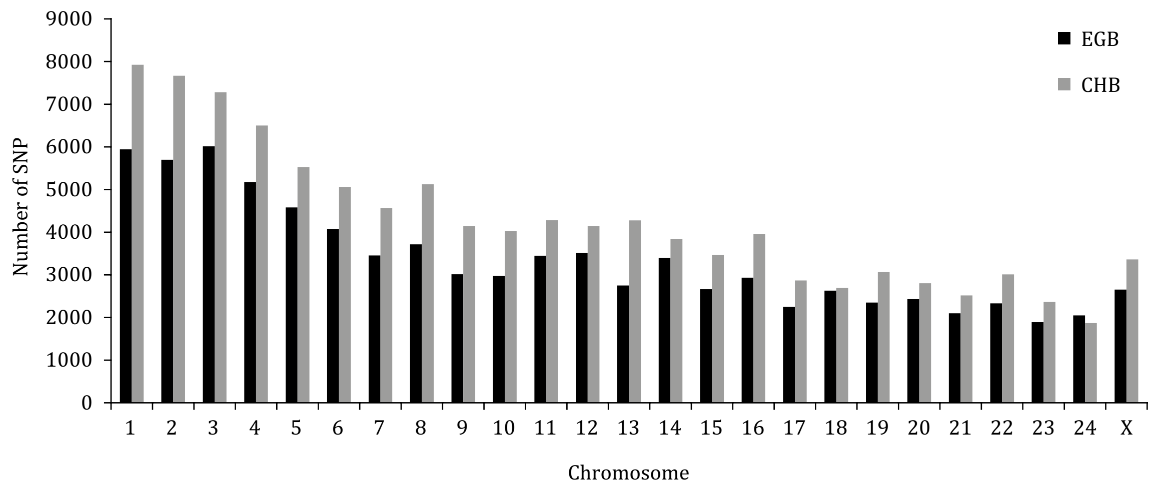


Figure 1 - Distribution of single nucleotide polymorphisms (SNP) on different chromosomes.

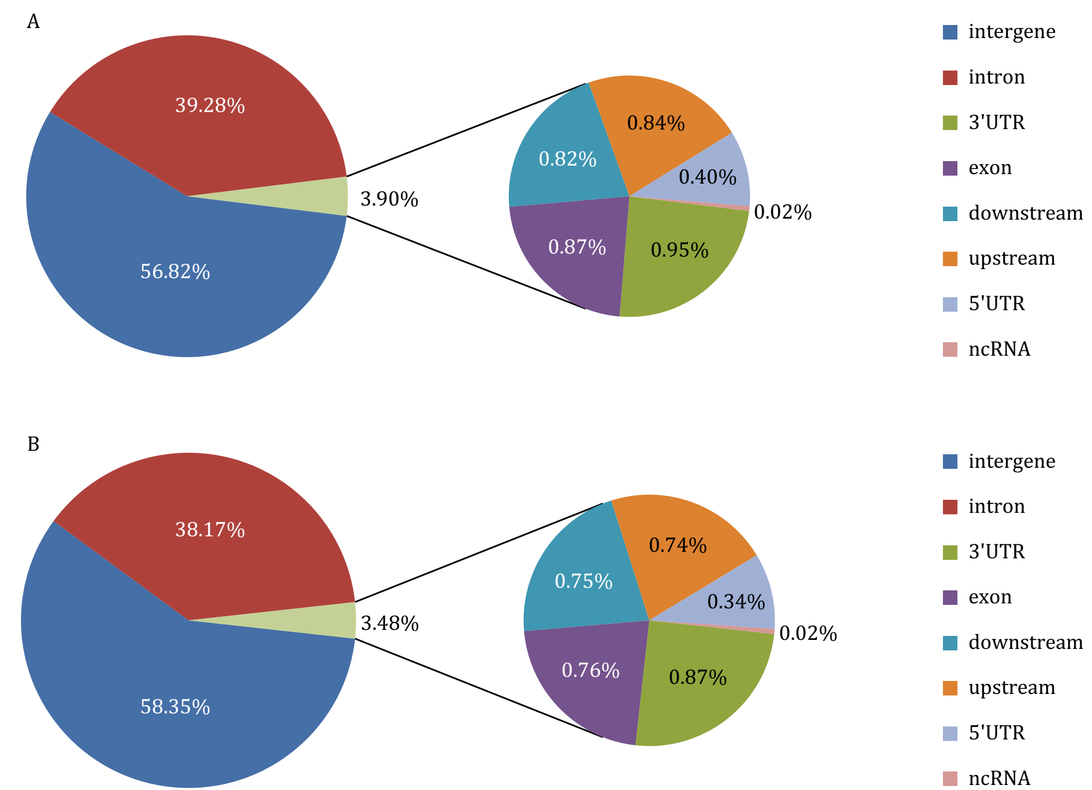
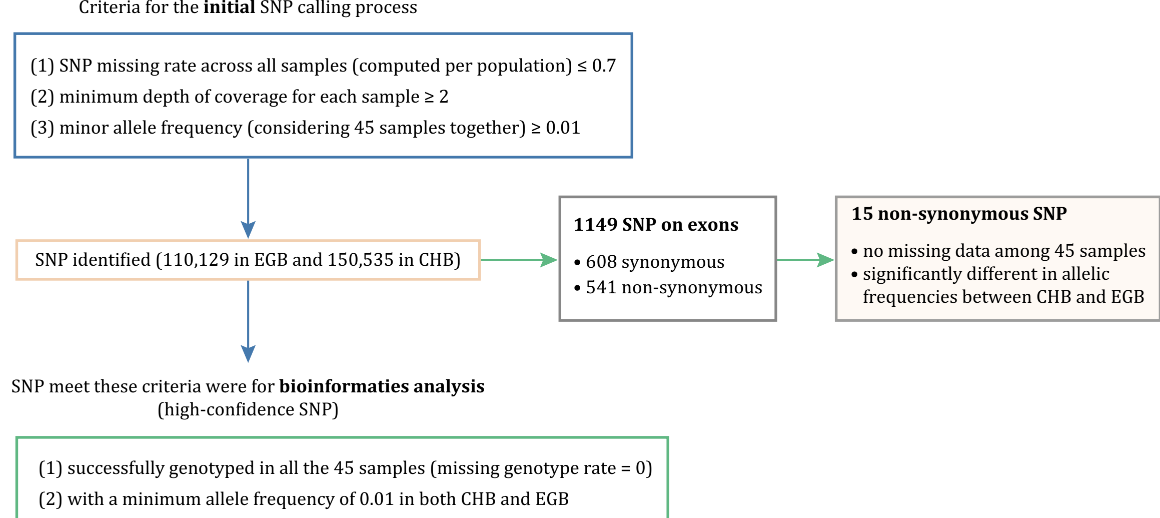


Figure 2 - Distribution of single nucleotide polymorphisms in the genome of the Egyptian (2a) and Chinese (2b) buffaloes.

differences (P<0.05) in the allelic frequencies between CHB and EGB, which yielded 15 SNP distributing on 12 chromosomes and resulting in non-synonymous amino acid mutations (Table 2).

For these 15 loci, the average reads coverage depth was 15X for EGB (ranging from 9 to 22X) and 20X for CHB (ranging from 13 to 28X). Except for the two SNP on *MTMR10* and *TACC2*, most polymorphisms were displayed only in one population. Furthermore, three mutations harbored by *RASGRP1* (K7E),



CHB - Chinese buffalo; EGB - Egyptian buffalo.

Figure 3 - Flow chart for single nucleotide polymorphism (SNP) detection and their further bioinformatics analysis.

Table 2 - Single nucleotide polymorphisms causing non-synonymous mutations and showing significant different allelic frequencies between two buffalo populations

Reference sequence (Chromosome)	Gene	Position in reference	Exon	Reference → Alteration		Allele frequency (Reference/alteration)		P-value
				Codon	Amino acid	CHB (n = 25)	EGB (n = 20)	
XM_006050687.2 (1)	*FAM90A26*	380	3	CTG → CCG	Leu → Pro	1.00/0.00	0.90/0.10	0.033
XM_006057766.2 (2)	*NEDD9*	421	2	ACT → GCT	Thr → Ala	1.00/0.00	0.83/0.17	0.002
XM_025279334.1 (3)	*CEP112*	2579	22	TCT → TTT	Ser → Phe	1.00/0.00	0.78/0.22	0.000
XM_025281542.1 (3)	*TPD52L3*	2440	1	GAC → GGC	Asp → Gly	1.00/0.00	0.83/0.17	0.003
XM_006052150.2 (6)	*CLCC1*	1752	11	ACG → ATG	Thr → Met	1.00/0.00	0.10/0.90	0.035
XM_006063939.2 (7)	*FAM184B*	2331	9	GAC → AAC	Asp → Asn	0.64/0.36	1.00/0.00	0.000
XM_006046572.2 (11)	*PNMA1*	1031	1	CCG → CTG	Pro → Leu	1.00/0.00	0.90/0.10	0.035
XM_025295940.1 (11)	*RASGRP1*	1482	1	AAA → GAA	Lys → Glu	1.00/0.00	0.60/0.40	0.000
XM_006051359.2 (14)	*SPATA25*	1633	2	CGG → CCG	Arg → Pro	1.00/0.00	0.78/0.23	0.000
XM_025266585.1 (16)	*PPFIBP2*	1863	16	GCA → ACA	Ala → Thr	1.00/0.00	0.83/0.17	0.002
XM_025271460.1 (20)	*MTMR10*	400	3	TGG → TCG	Trp → Ser	0.14/0.86	0.65/0.35	0.000
XM_025271849.1 (20)	*OTUD7A*	447	4	CAC → CAG	His → Gln	1.00/0.00	0.78/0.22	0.000
XM_025272296.1 (21)	*SLC22A14*	3016	10	GCC → ACC	Ala → Thr	0.50/0.50	1.00/0.00	0.000
XM_025274028.1 (23)	*TACC2*	4719	5	GAG → AAG	Glu → Lys	0.22/0.78	0.93/0.07	0.000
XM_025275895.1 (X)	*CACNA1F*	5237	43	ACG → ATG	Thr → Met	0.00/1.00	0.75/0.25	0.000

CHB - Chinese buffalo; EGB - Egyptian buffalo; *FAM90A26* - family with sequence similarity 90 member A26; *NEDD9* - neural precursor cell expressed, developmentally downregulated 9; *CEP112* - centrosomal protein 112; *TPD52L3* - tumor protein D52 like 3; *CLCC1* - chloride channel CLIC like 1; *FAM184B* - family with sequence similarity 184 member B; *PNMA1* - PNMA family member 1; *RASGRP1* - RAS guanyl releasing protein 1; *SPATA25* - spermatogenesis associated 25; *PPFIBP2* - PPFIA binding protein 2; *MTMR10* - myotubularin related protein 10; *OTUD7A* - OTU deubiquitinase 7A; *SLC22A14* - solute carrier family 22 member 14; *TACC2* - transforming acidic coiled-coil containing protein 2; *CACNA1F* - calcium voltage-gated channel subunit alpha1 F.

OTUD7A (H20Q), and *TACC2* (E1523K) caused alterations between acidic and alkaline amino acids. These SNP are especially worthy of further validation in more populations.

Over the last decade, accumulated studies have identified candidate genes on buffalo chromosome 1, 2, 3, 8, 12, 14, 15, 22, and 23 closely related to milk yield by either genome wide association study or candidate gene approach. In the present study, we identified some novel SNP located on these candidate genes (Table 3), which might provide potential evidence for their roles in milk production.

For further analysis, we retained only high-confidence SNP with no missing data among all the 45 samples, having at least a minimum coverage depth of 2, and a minimum allele frequency of 0.01 in both EGB and CHB populations. The final data set contained 1490 SNP with average reads coverage of 17.66X (Table 4). The majority of them were located in intronic (47.92%) and intergenic (47.25%) regions, followed by 1.74% of them in the 3'UTR and 1.21% in exons. The rest were located in the upstream (0.74%), downstream (0.67%), and 5'UTR (0.47%), respectively. Of all the 1490 SNP, 886 SNP loci associated with 658 functional genes were of particular interest to us due to their significant differences in allelic frequencies between EGB and CHB ($P<0.05$).

Gene ontology analysis of these 658 genes with the PANTHER classification system (v.14.0) revealed the top five biological processes (BP), molecular functions (MF), and cellular components (CC) in which these functional genes were involved (Figure 4). The fold enrichment for BP, MF, and CC ranged from 3.61 to 9.81, 5.21 to 7.09, and 2.69 to 3.99, respectively.

Table 3 - Novel single nucleotide polymorphisms (SNP) identified on known candidate genes associated with milk yield in buffaloes

BBU	Gene	Associated trait	Novel SNP identified
1	*BDH1*	Expression patterns in mammary epithelial cells during lactation positively correlated with milk yield in Murrah buffaloes (Yadav et al., 2015)	An upstream *BDH1* SNP (G/A at position 116814955)
2	*INSIG2*	305-d milk yield in Murrah and crossbred breeds (Deng et al., 2016b)	
2	*STAT1*	305-d milk yield and protein percentage in crossbred buffaloes (Deng et al., 2016a)	Intronic SNP (C/T at position 132437140)
2	*PRL*	Milk yield, peak milk yield, and the contents of protein and fat in milk in Italian Mediterranean river buffalo (Li et al., 2017)	
3	*STAT5A*	Milk yield trait in the Binlangjiang buffalo (Ji et al., 2013)	
3	*SREBF1*	305-d milk yield in crossbred Chinese buffaloes (Deng et al., 2017)	
3	*ACACA*	Expression patterns in mammary epithelial cells during lactation negatively correlated with milk yield in Murrah buffaloes (Yadav et al., 2015)	Intronic SNP (G/A at position 49760383 and 49760609, respectively)
8	*LEP*	High milk yield in Egyptian buffaloes (Nasr et al., 2016)	
12	*LIPIN1*	Expression patterns in mammary epithelial cells during lactation positively correlated with milk yield in Murrah buffaloes (Yadav et al., 2015)	
14	*ACSS2*	Expression patterns in mammary epithelial cells during lactation positively correlated with milk yield in Murrah buffaloes (Yadav et al., 2015)	A 3'UTR *ACSS2* SNP (G/A at position 19375891)
14	*OXT*	Milk yield in Italian buffaloes (Pauciullo et al., 2012)	
15	*VPS13B*	Total milk yield, fat yield, and protein yield (Liu et al., 2018a)	Intronic SNP (C/T at position 17737368)
15	*RGS22*	Total milk yield, fat yield, and protein yield (Liu et al., 2018a)	Intronic SNP (C/T at position 17857096)
22	*MC4R*	Milk yield trait in the Binlangjiang buffalo (Deng et al., 2016c)	
23	*SCD*	Expression patterns in mammary epithelial cells during lactation positively correlated with milk yield in Murrah buffaloes (Yadav et al., 2015)	

BBU - buffalo chromosomes; *BDH1* - 3-hydroxybutyrate dehydrogenase; *INSIG2* - insulin-induced gene 2; *STAT1* - signal transducer and activator of transcription 1; *PRL* - prolactin; *STAT5A* - signal transducer and activator of transcription 5A; *SREBF1* - sterol regulatory element binding transcription factor 1; *ACACA* - acetyl-coA carboxylase alpha; *LEP* - leptin; *LIPIN1* - lipin1; *OXT* - oxytocin; ACSS2 - acyl-CoA synthetase short-chain family member 2; *VPS13B* - vacuolar protein sorting 13 homolog B; *RGS22* - regulator of G protein signaling 22; *MC4R* - melanocortin 4 receptor; *SCD* - steroyl-CoA desaturase.

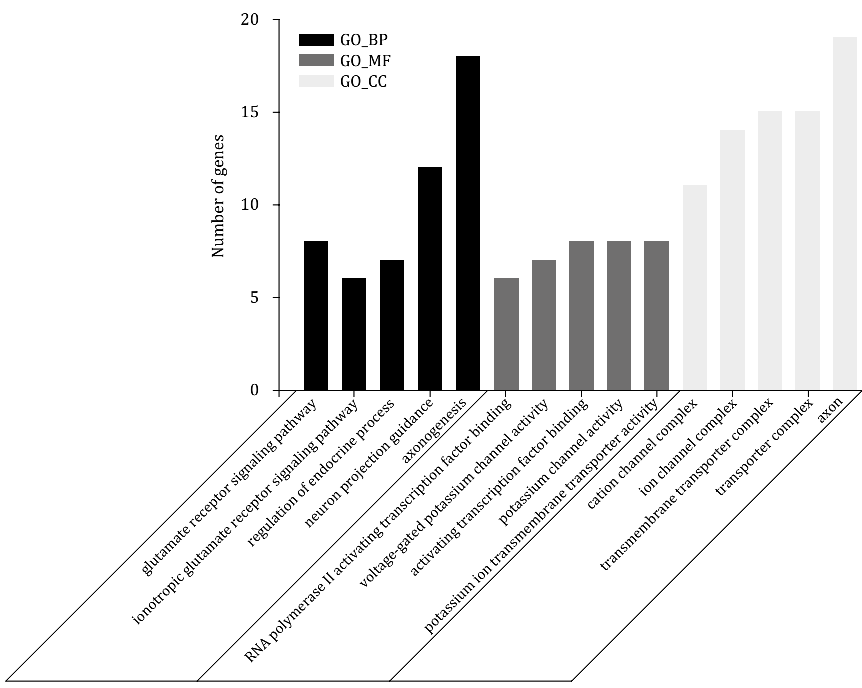
Panther pathways analysis showed that the most significantly over-represented pathways were the ionotropic glutamate receptor pathway (P = 1.18E-09, FDR = 1.94E-07), including eight genes (*CACNA1A*, *GRIA2*, *GRIA4*, *GRIK3*, *GRIK4*, *GRIK5*, *GRM3*, and *SHANK2*); the endothelin signaling pathway (P = 1.56E-04, FDR = 6.40E-03), including 10 genes (*ADCY2*, *ADCY5*, *EDNRA*, *PIK3R1*, *PIK3R6*, *PLCB1*, *PLCB4*, *PRKCE*, *PRKG1*, *PRKX*); and the gonadotropin-releasing hormone receptor pathway (P = 2.91E-05, FDR = 2.38E-03), including 14 genes (*ANXA5*, *CACNA1C*, *CREB1*, *FST*, *INHBA*, *INHBB*, *MAP3K1*, *NFATC2*, *PBX1*, *PIK3R1*, *PLCB1*, *PRKCE*, *SCG2*, *SMAD2*). The fold enrichment of three pathways was 7.00, 3.83, and 2.49, respectively. A total of 29 genes (Figure 5) were involved in these three pathways.

Further analysis revealed that nine (*ADCY5*, *CACNA1A*, *CREB1*, *INHBA*, *INHBB*, *PIK3R1*, *PLCB1*, *PRKCE*, and *SMAD2*) out of these 29 genes were significantly enriched in such biological processes as the regulation of hormone secretion, reproductive process, as well as ovulation (Figure 6). These genes, associated with high-confidence SNP showing significant differences in allelic frequencies between CHB and EGB, involved in the most significantly over-represented pathways, significantly enriched in the biological processes closely related to the activities of lactation, were considered to be the most interesting candidate genes for milk yield trait in buffaloes.

Table 4 - General information of high-confidence single nucleotide polymorphisms (SNP)

Chromosome	Number of SNP	Average coverage depth	
		CHB	EGB
1	132	20.01	15.86
2	127	20.09	15.38
3	96	18.67	15.30
4	81	19.18	15.67
5	101	19.14	16.11
6	58	20.16	15.15
7	70	20.70	15.68
8	66	19.69	14.52
9	52	17.73	15.57
10	51	19.66	14.53
11	44	20.54	14.81
12	74	19.95	17.44
13	34	17.33	14.54
14	51	19.35	15.89
15	41	20.05	14.90
16	49	20.63	16.44
17	45	19.85	15.68
18	41	19.50	15.92
19	48	20.34	14.84
20	44	19.72	16.02
21	37	19.82	15.22
22	45	19.43	16.87
23	46	19.24	14.93
24	40	21.30	16.84
X	17	19.96	16.95
Average	59.6	19.68	15.64

CHB - Chinese buffalo; EGB - Egyptian buffalo.



Functional genes referred to the 658 genes which were associated with high-confidence single nucleotide polymorphisms and differed significantly in allelic frequencies between Egyptian and Chinese buffaloes.

Figure 4 - Top five biological processes (BP), molecular functions (MF), and cell components (CC) in which functional genes were involved.

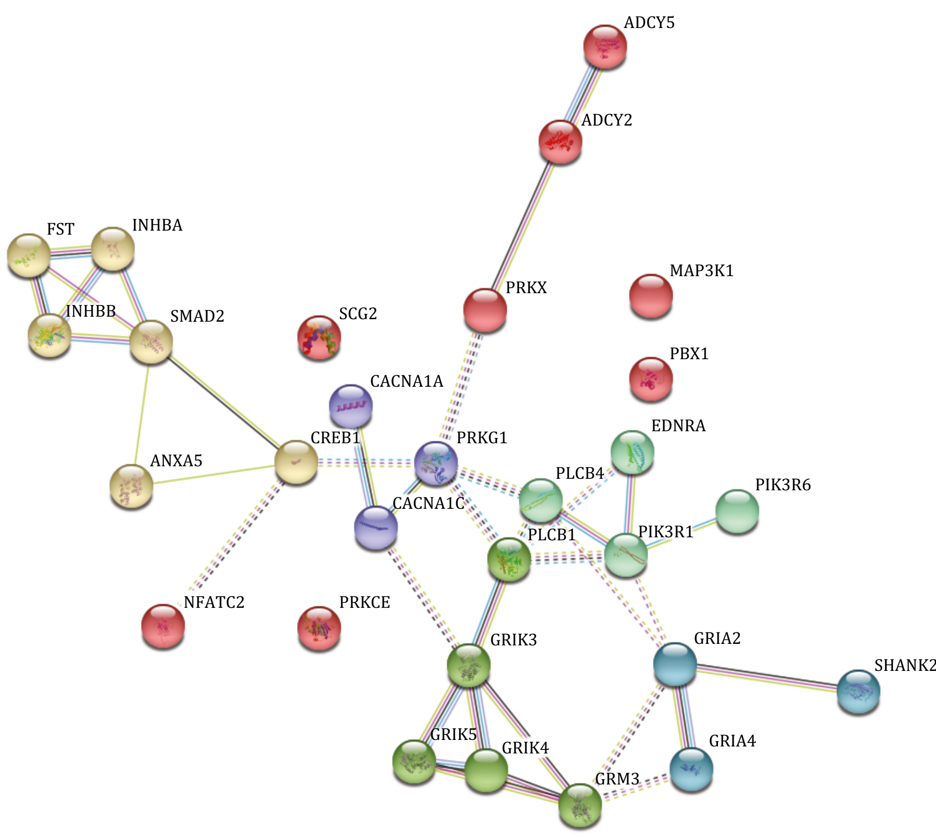


Figure 5 - Biological network of 29 genes involved in three significantly over-represented pathways (predicted by Strings, background: *Bos taurus*; number of clusters: 6).

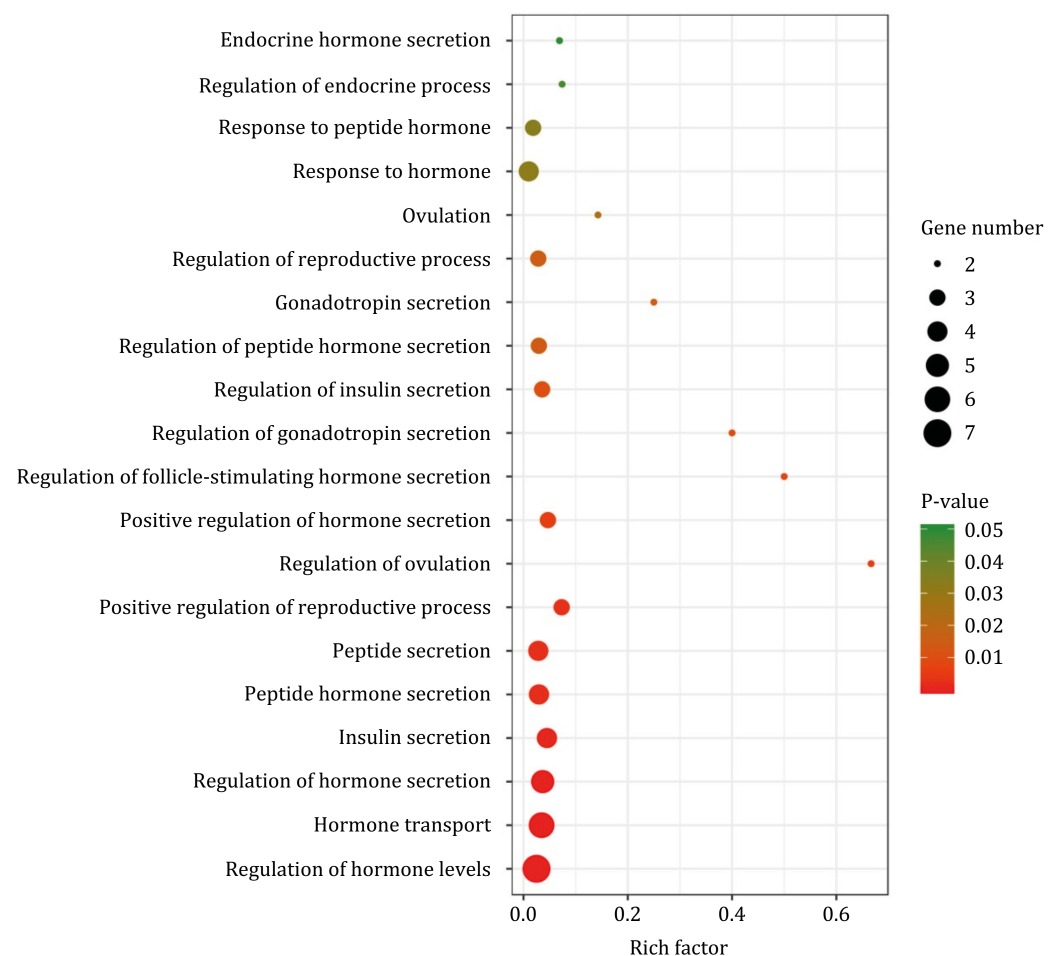


Figure 6 - Significantly enriched biological processes represented by nine candidate genes.

4. Discussion

Restriction site-associated DNA sequencing is a fast and useful technique to generate large numbers of SNP. Analyses based on SNP developed by RAD sequencing usually focused on high-confidence SNP only and removed those SNP showing low depth of coverage and low SNP calling rate among samples. The criteria set for screening of high-confidence SNP varied among documents. For example, a minimum coverage depth of 5 and more than 90% of the samples being successfully genotyped were adopted in a recent study (Gao et al., 2019). However, in another study, the minimum coverage depth was set to be 3, and 70% of the samples were required to have data to process a SNP (Hayashi et al., 2017). In the current study, a total of 886 high-confidence SNP were obtained. For each of them, the minimum coverage depth was 2, and the average reads coverage was 17.66X. All SNP were successfully genotyped in all the 45 samples with a minimum allele frequency of 0.01 and showed significant differences in allelic frequencies between CHB and EGB.

Functional genes associated with these high-confidence SNP were significantly over-represented in three pathways, which contained 29 genes with 41 related SNP. Of these, nine genes (*ADCY5*, *CACNA1A*, *CREB1*, *INHBA*, *INHBB*, *PIK3R1*, *PLCB1*, *PRKCE*, and *SMAD2*) were considered the most promising candidate genes for milk yield trait due to their modulating roles in hormonal regulation of the lactation cycle.

It is known that milk production is under the control of various hormones secreted by the neuroendocrine systems, which include the reproductive (such as estrogen, progesterone, prolactin, and oxytocin), metabolic (such as growth hormone, corticosteroids, thyroid hormones, and insulin),

as well as mammary (such as leptin) hormones (Neville et al., 2002; Svennersten-Sjaunja and Olsson, 2005; Crowley, 2015). Of all these nine genes, *INHBA*, *INHBB*, and *SMAD2* were significantly enriched in the TGF-β (transforming growth factor β) signaling pathway (P = 3.2E-3). The inhibin beta A subunit (INHBA) and inhibin beta B subunit (INHBB) are components of activins and inhibins, which belong to the TGF-β superfamily and are regulators for the synthesis and secretion of the pituitary follicle-stimulating hormone (FSH). Activins act as the stimulator. Inhibins, on the other hand, neutralize activins' activities by binding to them (Bilezikjian et al., 2006). They act with receptor-activated Smads (including Smad2) and co-mediator Smad (Smad4) to transfer the signal from cell surface to the nucleus and regulate the transcription of a variety of genes involved in follicular development, growth of oocytes (Knight and Glister, 2006; Xing et al., 2014), and embryo differentiation (Zhang et al., 2015). *ADCY5*, *CREB1*, and *PLCB1* were significantly enriched in the pathways of insulin secretion (P = 3.1E-3) and thyroid hormone synthesis (P = 2.2E-3). These three genes, together with *PIK3R1*, *PRKCE*, and *CACNA1A*, were also enriched in the pathways of estrogen signaling (P = 1.1E-4), aldosterone sythesis and secretion (P = 5.9E-5), and cholinergic synapse (P = 2.9E-6), respectively.

Due to their critical roles in hormone-related biological processes, association studies regarding intragenic SNP harbored by these functional genes and reproductive traits in farm animals have been conducted. For example, polymorphic loci in *INHBA* and *INHBB* were found to have significant effects on sperm quality and fertility in boars (Lin et al., 2006); intronic *INHBA* SNP was reported to be associated with fertility of stallions (Giesecke et al., 2010) and sperm quality in Chinese Holstein bulls (Sang et al., 2011); *INHBB* SNP (at 3'UTR) were related to litter size in sheep (Chu et al., 2011); *SMAD2* was considered an important candidate for total number born in swine (Wang et al., 2018). However, association studies regarding their sequencing variations and milk performance in buffaloes have not been documented yet. Based on the facts that these nine genes played various roles in the regulation of lactation-associated hormones, together with our findings that they were associated with high-confidence SNP showing significantly different allelic frequencies between CHB and EGB, we suggested that they were worthy of further investigations as candidate genes having influences on milk yield trait in buffaloes.

We also identified eight high-confidence SNP associated with five functional genes (*DIAPH3*, *FSTL4*, *GMDS*, *KCNMA1*, and *SLC44A5*), which have been previously documented as candidate genes for reproductive traits in buffaloes (Wu et al., 2013; de Camargo et al., 2015; Li et al., 2018a). They were all significantly different in the distribution of allele frequencies between EGB and CHB (Table 5). Five SNP

Table 5 - Single nucleotide polymorphisms (SNP) associated with known candidate genes closely related to reproductive traits in buffaloes

Region	Associated gene	BBU	Position	Reference (Ref)	Alteration (Alt)	Frequency of Ref allele	
						CHB	EGB
Intergenic	gene-*DIAPH3* (dist = 110981), gene-*TDRD3* (dist = 249397)	13	88461660	A	G	0.040	0.825
Intergenic	gene-*DIAPH3* (dist = 110637), gene-*TDRD3* (dist = 249741)	13	88461316	C	G	0.040	0.975
Intronic	gene-*FSTL4*	9	66157381	C	G	0.080	0.600
Intergenic	gene-*GMDS* (dist = 95662), gene-*MYLK4* (dist = 121469)	2	1453971	A	G	0.940	0.550
Intergenic	gene-*KCNMA1* (dist = 29841), gene-*DLG5* (dist = 108913)	4	152780961	T	C	0.920	0.600
Intergenic	gene-*KCNMA1* (dist = 30087), gene-*DLG5* (dist = 108667)	4	152781207	T	C	0.080	0.400
Intronic	gene-*SLC44A5*	6	68703492	A	G	0.980	0.350
Intronic	gene-*SLC44A5*	6	68703703	G	A	0.980	0.400

CHB - Chinese buffalo; EGB - Egyptian buffalo; BBU - buffalo chromosomes; *DIAPH3* - diaphanous-related formin 3; *TDRD3* - tudor-domain containing 3; *FSTL4* - follistatin-like 4; *GMDS* - GDP-mannose 4,6-dehydratase; *MYLK4* - myosin light chain kinase family member 4; *KCNMA1* - potassium calcium-activated channel subfamily M alpha 1; *DLG5* - discs large MAGUK scaffold protein 5; *SLC44A5* - solute carrier family 44 member 5.

were intergenic, such as SNP associated with *DIAPH3*, *GMDS*, and *KCNMA1*, which have been reported to be closely associated with calving interval and age at third calving in buffaloes. The remaining three SNP were intronic, with one and two SNP located within *FSTL4* and *SLC44A5*, respectively. In a previous study that combined GWAS (genome-wide association study) and RNA-seq of follicular granulosa cells, *FSTL4* was the nearest functional gene associated with a SNP closely related to age at second calving and age at third calving in Italian Mediterranean buffaloes (Li et al., 2018b). In the current study, a novel intronic SNP located on *FSTL4* was identified with significantly different allelic frequency distribution between CHB and EGB. Our results further confirmed that these functional genes, closely associated with genetic variants in two divergent populations with extreme phenotype in terms of milk yield, may explain the variances underlying the reproductive behavior in EGB and CHB.

In the present study, we used EGB and CHB, two phenotypically-divergent populations, to exploit genetic variants with potential effects on milk yield in buffaloes. Based on the strategy of selecting animals with extreme target trait for genotyping, previous association studies between DNA markers and interested traits have proved this to be an effective experimental design to identify candidate genes associated with target traits in other animals (Fontanesi et al., 2012a,b; Liu et al., 2018b). Together with the application of ddRAD-Seq, an efficient and cost-effective approach for SNP detection, genomic variants identified in this study provided additional insights into candidate genes affecting milk yield in buffaloes.

5. Conclusions

Our results provide potential genetic variances for the selection of milk yield trait in buffalo. We suggest nine genes which are involved in the hormonal regulation of lactation process as promising candidate genes worthy of further investigations for favorable alleles closely related to milk yield trait.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization: M. Ye and B. Zhou. Data curation: M. Ye, M. Xu and M. Lu. Formal analysis: M. Ye and M. Lu. Funding acquisition: B. Zhou. Methodology: M. Ye and M. Xu. Project administration: B. Zhou. Resources: M. Xu, M. Lu, H. Abd El-Kader, S.S. Alam and K.F. Mahrous. Software: M. Lu. Supervision: B. Zhou. Writing-original draft: M. Ye.

Acknowledgments

This work was supported by grants from Key Laboratory of Animal (Poultry) Genetics Breeding and Reproduction, Ministry of Agriculture and Rural Affairs (poultrylab2018-3) and an International Cooperation and Exchange program of the National Natural Science Foundation of China (31611140040). The authors thank the Seedstock Farm of Haizi Buffalo in Sheyang County, Yancheng, China for cooperation. We appreciated the great help provided by the head manager of the farm, Mr. Mingfa Gu, and the veterinary doctor, Mr. Liming Gu, in collecting the blood samples.

References

Bilezikjian, L. M.; Blount, A. L.; Donaldson, C. J. and Vale, W. W. 2006. Pituitary actions of ligands of the TGF-β family: activins and inhibins. Reproduction 132:207-215. https://doi.org/10.1530/rep.1.01073

Cheng, G. F.; Lu, A. H.; Liu, D. Q. and Gu, M. F. 2008. The conservation, exploitation and utilization of Haizi buffalo. China Cattle Science 34:60-62.

Chu, M.; Zhuang, H.; Zhang, Y.; Jin, M.; Li, X.; Di, R.; Cao, G.; Feng, T. and Fang, L. 2011. Polymorphism of inhibin β_B gene and its relationship with litter size in sheep. Animal Science Journal 82:57-61.

Cosenza, G.; Iannaccone, M.; Auzino, B.; Macciotta, N. P. P.; Kovitvadhi, A.; Nicolae, I. and Pauciullo, A. 2018. Remarkable genetic diversity detected at river buffalo *prolactin receptor* (*PRLR*) gene and association studies with milk fatty acid composition. Animal Genetics 49:159-168. https://doi.org/10.1111/age.12645

Cosenza, G.; Macciotta, N. P. P.; Nudda, A.; Coletta, A.; Ramunno, L. and Pauciullo, A. 2017. A novel polymorphism in the oxytocin receptor encoding gene (*OXTR*) affects milk fatty acid composition in Italian Mediterranean river buffalo. Journal of Dairy Research 84:170-180. https://doi.org/10.1017/S0022029917000127

Crowley, W. R. 2015. Neuroendocrine regulation of lactation and milk production. Comprehensive Physiology 5:255-291. https://doi.org/10.1002/cphy.c140029

da Costa Barros, C.; de Abreu Santos, D. J.; Aspilcueta-Borquis, R. R.; de Camargo, G. M. F.; de Araújo Neto, F. R. and Tonhati, H. 2018. Use of single-step genome-wide association studies for prospecting genomic regions related to milk production and milk quality of buffalo. Journal of Dairy Research 85:402-406. https://doi.org/10.1017/S0022029918000766

de Camargo, G. M. F.; Aspilcueta-Borquis, R. R.; Fortes, M. R. S.; Porto-Neto, R.; Cardoso, D. F.; Santos, D. J. A.; Lehnert, S. A.; Reverter, A.; Moore, S. S. and Tonhati, H. 2015. Prospecting major genes in dairy buffaloes. BMC Genomics 16:872. https://doi.org/10.1186/s12864-015-1986-2

Deng, T. X.; Pang, C. Y.; Lu, X. R.; Zhu, P.; Duan, A. Q. and Liang, X. W. 2016a. Associations between polymorphisms of the $STAT_1$ gene and milk production traits in water buffaloes. Journal of Animal Science 94:927-935. https://doi.org/10.2527/jas.2015-0139

Deng, T.; Pang, C.; Ma, X.; Lu, X.; Duan, A.; Zhu, P. and Liang, X. 2016b. Four novel polymorphisms of buffalo *INSIG2* gene are associated with milk production traits in Chinese buffaloes. Molecular and Cellular Probes 30:294-299. https://doi.org/10.1016/j.mcp.2016.09.003

Deng, T. X.; Pang, C. Y.; Liu, M. Q.; Zhang, C. and Liang, X. W. 2016c. Synonymous single nucleotide polymorphisms in the *MC4R* gene that are significantly associated with milk production traits in water buffaloes. Genetics and Molecular Research 15:gmr.15028153. https://doi.org/10.4238/gmr.15028153

Deng, T.; Pang, C.; Ma, X.; Duan, A.; Liang, S.; Lu, X. and Liang X. 2017. Buffalo *SREBP1*: molecular cloning, expression and association analysis with milk production traits. Animal Genetics 48:720-721. https://doi.org/10.1111/age.12587

Du, C.; Deng, T.; Zhou, Y.; Ye, T.; Zhou, Z.; Zhang, S.; Shao, B.; Wei, P.; Sun, H.; Khan, F. A.; Yang, L. and Hua, G. 2019. Systematic analyses for candidate genes of milk production traits in water buffalo (*Bubalus Bubalis*). Animal Genetics 50:207-216. https://doi.org/10.1111/age.12739

El-Halawany, N.; Abdel-Shafy, H.; Shawky, A. E. M. A.; Abdel-Latif, M. A.; Al-Tohamy, A. F. M. and Abd El-Moneim, O. M. 2017. Genome-wide association study for milk production in Egyptian buffalo. Livestock Science 198:10-16. https://doi.org/10.1016/j.livsci.2017.01.019

Fontanesi, L.; Bertolini, F.; Dall'Olio, S.; Buttazzoni, L.; Gallo, M. and Russo, V. 2012a. Analysis of association between the *MUC4* g.8227C>G polymorphism and production traits in Italian heavy pigs using a selective genotyping approach. Animal Biotechnology 23:147-155. https://doi.org/10.1080/10495398.2011.653462

Fontanesi, L.; Buttazzoni, L.; Scotti, E. and Russo, V. 2012b. Confirmation of the association between a single nucleotide polymorphism in the porcine *LDHA* gene and average daily gain and correlated traits in Italian Large White pigs. Animal Genetics 43:649-650. https://doi.org/10.1111/j.1365-2052.2012.02355.x

Gao, Y.; Yin, S.; Zhang, L.; Liao, Y. Y.; Guan, J. L. and Tang, L. Z. 2019. Development of 30 SNP markers for *Amorphophallus yunnanensis* based on RAD sequencing. Conservation Genetics Resources 11:423-426. https://doi.org/10.1007/s12686-018-1040-1

Giesecke, K.; Hamann, H.; Sieme, H. and Distl, O. 2010. *INHBA*-associated markers as candidates for stallion fertility. Reproduction in Domestic Animals 45:342-347. https://doi.org/10.1111/j.1439-0531.2008.01325.x

Gu, M.; Cosenza, G.; Iannaccone, M.; Macciotta, N. P. P.; Guo, Y.; Di Stasio, L. and Pauciullo, A. 2019. The single nucleotide polymorphism g.133A>C in the stearoyl CoA desaturase gene (*SCD*) promoter affects gene expression and quali-quantitative properties of river buffalo milk. Journal of Dairy Science 102:442-451. https://doi.org/10.3168/jds.2018-15059

Hayashi, Y.; Oguchi, K.; Yamaguchi, K.; Kitade, O.; Maekawa, K.; Miura, T. and Shigenobu, S. 2017. Male-specific molecular genetic markers in the Japanese subterranean termite *Reticulitermes speratus*. Insectes Sociaux 64:357-364. https://doi.org/10.1007/s00040-017-0553-z

Iamartino, D.; Nicolazzi, E. L.; Van Tassell, C. P.; Reecy, J. M.; Fritz-Waters, E. R.; Koltes, J. E.; Biffani, S.; Sonstegard, T. S.; Schroeder, S. G.; Ajmone-Marsan, P.; Negrini, R.; Pasquariello, R.; Ramelli, P.; Coletta, A.; Garcia, J. F.; Ali, A.; Ramunno, L.; Cosenza, G.; Oliveira, D. A. A.; Drummond, M. G.; Bastianetto, E.; Davassi, A.; Pirani, A.; Brew, F. and Williams, J. L. 2017. Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). PLoS One 12:e0185220. https://doi.org/10.1371/journal.pone.0185220

Ji, M.; Liu, X. H.; Yu, C. L.; Yu, X. F.; Liu, D. J. and Shi, X. W. 2013. Association of *STAT5A* polymorphism and milk production traits in Binlangjiang buffalo. China Cattle Science 39:29-35.

Knight, P. G. and Glister C. 2006. TGF-β superfamily members and ovarian follicle development. Reproduction 132:191-206. https://doi.org/10.1530/rep.1.01074

Li, J.; Liang, A.; Li, Z.; Du, C.; Hua, G.; Salzano, A.; Campanile, G.; Gasparrini, B. and Yang, L. 2017. An association analysis between *PRL* genotype and milk production traits in Italian Mediterranean river buffalo. Journal of Dairy Research 84:430-433. https://doi.org/10.1017/S0022029917000693

Li, J.; Liu, J. J.; Campanile, G.; Plastow, G.; Zhang, C. Y.; Wang, Z. Q.; Cassandro, M.; Gasparrini, B.; Salzano, A.; Hua, G. H.; Liang, A. X. and Yang, L. G. 2018a. Novel insights into the genetic basis of buffalo reproductive performance. BMC Genomics 19:814. https://doi.org/10.1186/s12864-018-5208-6

Li, J.; Liu, J.; Liu, S.; Plastow, G.; Zhang, C.; Wang, Z.; Campanile, G.; Salzano, A.; Gasparrini, B.; Hua, G.; Liang, A. and Yang, L. 2018b. Integrating RNA-seq and GWAS reveals novel genetic mutations for buffalo reproductive traits. Animal Reproduction Science 197:290-295. https://doi.org/10.1016/j.anireprosci.2018.08.041

Li, J.; Liu, S.; Li, Z.; Zhang, S.; Hua, G.; Salzano, A.; Campanile, G.; Gasparrini, B.; Liang, A. and Yang, L. 2018c. *DGAT1* polymorphism in Riverine buffalo, Swamp buffalo and crossbred buffalo. Journal of Dairy Research 85:412-415. https://doi.org/10.1017/S0022029918000468

Lin, C. L.; Ponsuksili, S.; Tholen, E.; Jennen, D. G. J.; Schellander, K. and Wimmers, K. 2006. Candidate gene markers for sperm quality and fertility of boar. Animal Reproduction Science 92:349-363. https://doi.org/10.1016/j.anireprosci.2005.05.023

Liu, J. J.; Liang, A. X.; Campanile, G.; Plastow, G.; Zhang, C.; Wang, Z.; Salzano, A.; Gasparrini, B.; Cassandro, M. and Yang, L. G. 2018a. Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. Journal of Dairy Science 101:433-444. https://doi.org/10.3168/jds.2017-13246

Liu, J.; Liu, R. R.; Wang, J.; Zhang, Y. H.; Xing, S. Y.; Zheng, M. Q.; Cui, H. X.; Li, Q. H.; Li, P.; Cui, X. Y.; Li, W.; Zhao, G. P. and Wen, J. 2018b. Exploring genomic variants related to residual feed intake in local and commercial chickens by whole genomic resequencing. Genes 9:57. https://doi.org/10.3390/genes9020057

Manzoor, S.; Nadeem, A. and Javed, M. 2020. Polymorphism association and expression analysis of alpha-lactalbumin (*LALBA*) gene during lactation in Nili Ravi buffalo. Tropical Animal Health and Production 52:265-271. https://doi.org/10.1007/s11250-019-02010-0

Manzoor, S.; Nadeem, A.; Maryam, J.; Hashmi, A. S.; Imran, M. and Babar, M. E. 2018. Osteopontin gene polymorphism association with milk traits and its expression analysis in milk of riverine buffalo. Tropical Animal Health and Production 50:275-281. https://doi.org/10.1007/s11250-017-1426-1

Menon, R.; Patel, A. B. and Joshi, C. 2016. Comparative analysis of SNP candidates in disparate milk yielding river buffaloes using targeted sequencing. PeerJ 4:e2147. https://doi.org/10.7717/peerj.2147

Michelizzi, V. N.; Dodson, M. V.; Pan, Z. X.; Amaral, M. E. J.; Michal, J. J.; McLean, D. J.; Womack, J. E. and Jiang, Z. H. 2010. Water buffalo genome science comes of age. International Journal of Biological Sciences 6:333-349. https://doi.org/10.7150/ijbs.6.333

Mi, H.; Muruganujan, A.; Ebert, D.; Huang, X. and Thomas, P. D. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research 47:D419-D426. https://doi.org/10.1093/nar/gky1038

Mi, H. and Thomas, P. 2009. PANTHER Pathway: an ontology-based pathway database coupled with data analysis tools. Methods in Molecular Biology 563:123-140.

Nasr, M. A. F.; Awad, A. and El Araby, I. E. 2016. Associations of *leptin* and *pituitary-specific transcription factor* genes' polymorphisms with reproduction and production traits in dairy buffalo. Reproduction in Domestic Animals 51:597-603. https://doi.org/10.1111/rda.12726

Neville, M. C.; McFadden, T. B. and Forsyth, I. 2002. Hormonal regulation of mammary differentiation and milk secretion. Journal of Mammary Gland Biology and Neoplasia 7:49-66. https://doi.org/10.1023/A:1015770423167

Pauciullo, A.; Cosenza, G.; Steri, R.; Coletta, A.; Jemma, L.; Feligini, M.; Di Berardino, D.; Macciotta, N. P. and Ramunno, L. 2012. An association analysis between *OX*T genotype and milk yield and flow in Italian Mediterranean river buffalo. Journal of Dairy Research 79:150-156. https://doi.org/10.1017/s0022029911000914

Peterson, B. K.; Weber, J. N.; Kay, E. H.; Fisher, H. S. and Hoekstra, H. E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Sang, L.; Du, Q. Z.; Yang, W. C.; Tang, K. Q.; Yu, J. N.; Hua, G. H.; Zhang, X. X. and Yang, L. G. 2011. Polymorphisms in follicle stimulation hormone receptor, inhibin alpha, inhibin bata A, and prolactin genes, and their association with sperm quality in Chinese Holstein bulls. Animal Reproduction Science 126:151-156. https://doi.org/10.1016/j.anireprosci.2011.04.023

Shi, D. S.; Wang, J.; Yang, Y.; Lu, F. H.; Li, X. P. and Liu, Q. Y. 2012. *DGAT1*, *GH*, *GHR*, *PRL* and *PRLR* polymorphism in water buffalo (*Bubalus bubalis*). Reproduction in Domestic Animals 47:328-334. https://doi.org/10.1111/j.1439-0531.2011.01876.x

Svennersten-Sjaunja, K. and Olsson, K. 2005. Endocrinology of milk production. Domestic Animal Endocrinology 29:241-258. https://doi.org/10.1016/j.domaniend.2005.03.006

Venturini, G. C.; Cardoso, D. F.; Baldi, F.; Freitas, A. C.; Aspilcueta-Borquis, R. R.; Santos, D. J. A.; Camargo, G. M. F.; Stafuzza, N. B.; Albuquerque, L. G. and Tonhati, H. 2014. Association between single-nucleotide polymorphisms and milk production traits in buffalo. Genetics and Molecular Research 13:10256-10268. https://doi.org/10.4238/2014.december.4.20

Wang, Y.; Ding, X.; Tan, Z.; Xing, K.; Yang, T.; Pan, Y.; Wang, Y.; Mi, S.; Sun, D. and Wang, C. 2018. Genome-wide association study for reproductive traits in a Large White pig population. Animal Genetics 49:127-131. https://doi.org/10.1111/age.12638

Wu, J. J.; Song, L. J.; Wu, F. J.; Liang, X. W.; Yang, B. Z.; Wathes, D. C.; Pollott, G. E.; Cheng, Z.; Shi, D. S.; Liu, Q. Y.; Yang, L. G. and Zhang, S. J. 2013. Investigation of transferability of BovineSNP50 BeadChip from cattle to water buffalo for genome wide association study. Molecular Biology Reports 40:743-750. https://doi.org/10.1007/s11033-012-1932-1

Xing, N.; Liang, Y.; Gao, Z.; He, J.; He, X.; Li, H. and Dong, C. 2014. Expression and localization of Smad2 and Smad4 proteins in the porcine ovary. Acta Histochemica 116:1301-1306. https://doi.org/10.1016/j.acthis.2014.07.014

Yadav, P.; Kumar, P.; Mukesh, M.; Kataria, R. S.; Yadav, A.; Mohanty, A. K. and Mishra, B. P. 2015. Kinetics of lipogenic genes expression in milk purified mammary epithelial cells (MEC) across lactation and their correlation with milk and fat yield in buffalo. Research in Veterinary Science 99:129-136. https://doi.org/10.1016/j.rvsc.2015.01.003

Zhang, K.; Rajput, S. K.; Lee, K. B.; Wang, D.; Huang, J.; Folger, J. K.; Knott, J. G.; Zhang, J. Z. and Smith, G. W. 2015. Evidence supporting a role for SMAD2/3 in bovine early embryonic development: potential implications for embryotropic actions of follistatin. Biology of Reproduction 93:86. https://doi.org/10.1095/biolreprod.115.130278

Zhu, J. C.; Guo, Y. S.; Su, K.; Liu, Z. D.; Ren, Z. H.; Li, K. and Guo, X. W. 2018. Construction of a highly saturated genetic map for *Vitis* by next-generation restriction site-associated DNA sequencing. BMC Plant Biology 18:347. https://doi.org/10.1186/s12870-018-1575-z