Journal of Dairy Research

cambridge.org/dar

Prospecting genomic regions associated with milk production traits in Egyptian buffalo

Hamdy Abdel-Shafy[1], Mohamed A. A. Awad[1], Hussein El-Regalaty[2], S. E.-D. El-Assal[3] and Samy Abou-Bakr[1]

[1]Department of Animal Production, Faculty of Agriculture, Cairo University, El-Gamma street, 12613 Giza, Egypt; [2]Department of buffalo research, Animal Production Research Institute, Agricultural Research Center, Dokki, Giza, Egypt and [3]Department of Genetics, Faculty of Agriculture, Cairo University, El-Gamma street, 12613 Giza, Egypt

Research Article

Cite this article: Abdel-Shafy H, Awad MAA, El-Regalaty H, El-Assal SE-D and Abou-Bakr S (2020). Prospecting genomic regions associated with milk production traits in Egyptian buffalo. *Journal of Dairy Research* **87**, 389–396. https://doi.org/10.1017/S0022029920000953

Received: 4 March 2020
Revised: 3 August 2020
Accepted: 10 August 2020
First published online: 13 November 2020

Keywords:
Candidate gene; genetic markers; genome-wide association; milk quality

Author for correspondence:
Hamdy Abdel-Shafy,
Email: hamdyabdelshafy@agr.cu.edu.eg

Abstract

The objectives of the current study were to detect putative genomic loci and to identify candidate genes associated with milk production traits in Egyptian buffalo. A total number of 161 479 daily milk yield (DMY) records and 60 318 monthly measures for fat and protein percentages (FP and PP, respectively), along with fat and protein yields (FY and PY, respectively) from 1670 animals were used. Genotyping was performed using Axiom® Buffalo Genotyping 90 K array. Genome-wide association study (GWAS) for each trait was performed using PLINK. After Bonferroni correction, 47 SNPs were associated with one or more milk production traits. These SNPs were distributed over 36 quantitative trait loci (QTL) and located on 20 buffalo chromosomes (BBU). For the 47 SNPs, one was overlapped for three traits (DMY, FY, and PY), six were associated with two traits (one for PP and PY and five for FY and PY) while the rest were associated with only one trait. Out of 36 identified QTL, eleven were overlapped with previously reported loci in buffalo and/or cattle populations. Some of these SNPs are placed within or close to potential candidate genes, for example: *TPD52, ZBTB10, RALYL* and *SNX16* on BBU15, *ADGRD1* on BBU17, *ESRRG* on BBU5 and *GRIP1* on BBU4. This is the first reported study between genome-wide markers and milk components in Egyptian buffalo. Our findings provide useful information to explore the genetic mechanisms and relevant genes contributing to the variation in milk production traits. Further confirmation studies with larger population size are necessary to validate the findings and detect the causal genetic variants.

Egyptian buffalo have played an important role over centuries in securing food supply (milk and meat) and providing manure for agricultural operations in rural regions (SADS, 2009). In Egypt, there are around 3.4 million buffalo producing around 45 and 37% of total milk and red meat production, respectively (FAO, 2019). Buffalo milk has around 22% more protein, 47% more fat, 50% more iron and 50% less lactose as well as lower cholesterol content as compared with typical bovine milk (Barłowska *et al*., 2011). In addition, buffalo are more adapted to harsh environmental conditions than other dairy species. Likewise, buffalo are better able to utilize poor quality roughages, are more resistant to several infectious diseases and have longer productive life than cows (Wanapat *et al*., 2000; Deb *et al*., 2016). Although considerable attention has been paid in the last decades for buffalo to achieve their production potential, genetic improvement for buffalo performance is still not satisfactory. The main reason for this delay is the lack of national recording systems and pedigree information. To partially overcome this problem, an intensive crossing with Italian buffalo was applied by the traditional breeding industry. Although significant progress has been achieved, milk yield is still considerably lower than potentially expected. An alternative and potentially more rapid approach is to use genomic information in breeding programs. Genomic data can identify essential molecular markers and biological pathways to better understand the genetic mechanisms underlying milk properties so as to accelerate breeding progress (Hayes and Daetwyler, 2019).

Advances in genomic technologies have led to the identification of many thousands of genomic markers covering the whole genome, mainly on a single base (single nucleotide polymorphisms: SNPs). Consequently, genome-wide association studies (GWAS) have been widely accepted as a primary approach for localizing quantitative trait loci (QTL) and have achieved considerable success with detecting candidate genes associated with economically complex traits in livestock species and disease risks in human (Goddard and Hayes, 2009; Visscher *et al*., 2017). For buffalo, development of Axiom Buffalo Genotyping Array (90k SNPs) opens new opportunities to identify candidate genes associated with buffalo milk production traits and provides the possibility of predicting genomic breeding values for buffalo breeds in different countries. So far, 87 significant and 354 suggestive genomic regions were reported for milk production traits in Brazilian, Chinese, Egyptian, Iranian, Italian and

Hannah Dairy Research Foundation

CAMBRIDGE UNIVERSITY PRESS

Philippino buffalo (de Camargo *et al.*, 2015; El-Halawany *et al.*, 2017; Iamartino *et al.*, 2017; Mokhber, 2017; da Costa Barros *et al.*, 2018; Herrera *et al.*, 2018; Liu *et al.*, 2018; Abdel-Shafy *et al.*, 2020; Awad *et al.*, 2020; Lu *et al.*, 2020). Despite the fact that these loci were distributed on almost all chromosomes, none of these SNPs were overlapped between studies indicating the complexity of milk production traits and the modest effect of such SNPs (Abdel-Shafy *et al.*, 2020). Therefore, a number of QTL and novel genes underlying buffalo milk properties may still need to be identified and/or validated. Motivated by these reasons, the objectives of the current study were to identify genomic regions and detect potential candidate genes associated with milk production traits in Egyptian buffalo.

Materials and methods

The study was approved by the Institutional Animal Care and Use Committee (IACUC) of Cairo University, Egypt (approval number: CU-II-F-40-17).

Animals and phenotypes

Data consisted of 177 995 daily milk records from 1992 multiparous Egyptian buffalo cows belonging to eleven herds. To ensure homogenous dataset, a number of quality criteria were applied to animals and phenotypes (online Supplementary Table S1). Animals with unknown calving dates, records of lactation number above thirteen, and data below five and above 290 days in milk (DIM) were excluded. Furthermore, animal must have completed at least three months of lactation to be used in the analyses. After quality check, 161 479 daily milk records from 1670 animals remained. The average lactation length during the first thirteen lactations was 199 ± 54 d (mean ± SD). The pedigree file included 2121 animals. The average daily milk yield (DMY) in the filtered data was 7.8 ± 3.0 kg. All information was obtained from Cattle Information System/Egypt (CISE) and Animal Production Research Institute (APRI) of Agricultural Research Center (ARC), Egypt.

Milk samples from the above filtered animals were monthly collected and stored at −20°C until analysis. Time between subsequent milk samples for the same animal was 31 ± 9 d. The total number of analyzed samples was 60 318 from 1481 animals. Fat (FP), and protein (PP) percentages were determined by Infrared Milk Analyzer (Bentley). The average of FP and PP were 7 ± 2.0 and 3.7 ± 0.6, respectively. Fat (FY), and protein (PY) yields were calculated by multiplying percentages by milk yield at same test day. The averages of FY and PY were 0.62 ± 0.30 and 0.33 ± 0.13 kg/day, respectively.

Genotypic data

A total number of 114 animals were selected for genotyping according to their average daily milk, where animals with highest and lowest deviation from the population mean were used. Blood samples were collected from the jugular vein and kept in a 15 ml Falconer tubes containing 1 ml 0.5 M EDTA and chilled immediately on ice. Genomic DNA was extracted from whole blood samples using QIAamp® DNA Blood Mini Kit (QIAGEN, Hilden, Germany). Genotyping was performed with Axiom® Buffalo Genotyping panel 90 K according to the standard protocol of Thermo Fisher Scientific (https://www.thermofisher.com). The raw signal intensities files (CEL format) were converted into genotype calls and annotated to the latest reference assembly of buffalo genome (UOA_WB_1: GCA_003121395.1) using Genotyping Console™ 4.2.

Statistical analysis

A yield deviation (adjusted phenotype) is a weighted average of the animal's phenotype adjusted for non-genetic factors (VanRaden and Wiggans, 1991). In this regard, yield deviation for each trait was computed with univariate animal model using BLUPF90 family (Misztal *et al.*, 2002). Given the vector of y representing the phenotypic observations on the tested trait (DMY, FP, PP, FY, and PY), the following univariate animal model was used: $y = Xb + Z\alpha + Wp + \varepsilon$, where b is the vector of all fixed effects including milking frequency per day, lactation number, herd, and year-season of calving. In addition, linear regressions of age at calving, and fourth order Legendre polynomials of DIM were used. While α and p are the vectors of random additive genetic and permanent environmental effects, respectively, and ε is the vector of random residual. X, Z and W are incidence matrices relating observations of y to fixed, random animal, and random permanent environmental effects, respectively. The analysis was performed under the following assumptions: $\propto \sim N(0, A\sigma_a^2)$, $p \sim N(0, I\sigma_p^2)$, and $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$, where A is a matrix of additive genetic relationship among animals based on available pedigree; I is an identity matrix; and $\sigma^2\alpha$, σ^2p and $\sigma^2\varepsilon$ are additive, permanent environmental and residual variances, respectively.

The GWAS was performed using the linear regression model implemented in PLINK 1.9 (Chang *et al.*, 2015), where the adjusted phenotype (DMY, FP, PP, FY, and PY) were regressed on the number of copies of the alleles using PLINK–linear option with population stratification as covariates. In this issue, a multidimensional scaling approach was used to adjust for a potential population structure. The scaling process identified eight significant clusters indicating the axes of ancestry at $P < 0.0001$. These clusters were included as covariates into the model when performing GWAS. The Manhattan and Q-Q plots were generated using qqman package in R (Turner, 2014). For statistical inference, the Bonferroni method was applied to correct for multiple testing.

QTL and candidate genes

Beside previously reported loci in buffalo studies, reported QTL for milk production traits in cattle were retrieved from animal QTLdb (http://www.animalgenome.org/QTLdb), release 39 (Hu *et al.*, 2019), since buffalo and cattle are closely related. Candidate genes in each genomic region were extracted from the latest annotated file (na35.r2.a2) for Axiom® Buffalo Genotyping array provided by Thermo Fisher Scientific. Gene functions are extracted from UniProtKB (https://www.uniprot.org/) and GeneCards (https://www.genecards.org/) databases.

Results and discussion

Animals and phenotypes

To accurately detect genomic loci associated with complex traits, it is crucial to use a suitable number of individuals with accurate phenotypes (Goddard and Hayes, 2009). The number of animals used in current GWAS is relatively small compared to other investigations performed in dairy cattle populations. The main limiting

https://doi.org/10.1017/S0022029920000953 Published online by Cambridge University Press

reason for increasing the sample size is the availability of accurate daily milk records, since it is not routinely recorded in buffalo farms. In addition, there is currently an intensive crossing with Italian buffalo in most of the commercial herds, which makes it difficult to find large scale data from pure Egyptian buffalo. To circumvent such a situation, we used data from institutional herds, where pure Egyptian buffalo are kept. In addition, we firstly corrected phenotypes for non-genetic factors using 161 479 daily milk records and 60 318 milk composition records from 1670 animals. We reduced the heterogeneity among animals and lactations as much as possible. In this regard, the coefficients of variation among lactations were 0.41, 0.27, 0.16, 0.43, and 0.36 for DMY, FP, PP, FY, and PY, respectively. Adjusted phenotypes (yield deviations) are commonly used when assessing the animal's own performance to reduce variability attributed to non-genetic factors, increase the accuracy of estimating the genetic effect and to reduce the error rate (VanRaden and Wiggans, 1991). Later, we used adjusted phenotypes for GWAS analyses.

Genotype characteristics

In the most recent version of annotation file (na35.r2.a2) provided by Thermo Fisher Scientific, the number of chromosomes is 25 including the sex chromosomes. The previous version of assembly (r1) used in the previous investigations was annotated to bovine genome (UMD31 assembly: see de Camargo *et al.* (2015); Iamartino *et al.* (2017); da Costa Barros *et al.* (2018); Liu *et al.* (2018); Herrera *et al.* (2018); El-Halawany *et al.* (2017); Mokhber (2017)). Therefore, the number of chromosomes in the previous versions of this array was 30. The current version of the Axiom Buffalo Genotyping array comprises 123 040 SNPs. The average probe space between these markers in relation to the buffalo UOA_WB_1 assembly was one SNP every 34.46 kb across all loci.

In the current study, the raw genotypic data were subjected to quality control procedures performed with PLINK 1.9 (Chang *et al.*, 2015). In this respect; almost 48% from the genotyped SNPs was omitted due to unknown or duplicated positions, missing genotype per a SNP >0.15, low minor frequency (MAF) <0.01, and/or significant ($P < 0.0001$) deviation from Hardy-Weinberg proportion (online Supplementary Table S2). Individuals with low call rate (>0.15) were also discarded. After adopting the filtering options, the number of genotyped animals and SNPs were 113 and 64 169, respectively. The genotyping rate for the remaining animals was 98.4%. To reduce the redundant information and to prevent bias that could skew the association tests for GWAS (Anderson *et al.*, 2010; Laurie *et al.*, 2010), we excluded one of a pair of SNPs if the LD of $r^2 > 0.5$ within a sliding window of 50 SNPs and moving 5 SNPs per set using PLINK (Chang *et al.*, 2015).This LD pruning option step led to reduce the SNP number to 44 985 markers.

The filtered SNPs covered ~2.6 Gb of the buffalo genome with average distance of 40.8 ± 32.0 kb between SNPs (Table 1) and average MAF of 0.29 ± 0.13 across chromosomes. While in the previous investigation in Egyptian buffalo, the average density and MAF were 62.66 ± 67.10 and 0.30 ± 0.13, respectively (El-Halawany *et al.*, 2017). Because most of SNP markers are outside and/or away from the candidate genes, the lower distance would lead us to detect genomic loci more accurately and facilitate detection of candidate genes (Brodie *et al.*, 2016). The square of correlation coefficient (r^2) is commonly used to determine the level of LD between markers. It has been suggested that GWAS and QTL mapping are better performed when $r^2 \geq 0.3$ (Ardlie *et al.*, 2002). In the current study, the average r^2 between adjacent markers is equal to 0.43 ± 0.21 providing powerful information for performing GWAS.

Genome-wide association analyses

An important issue in GWAS is to ensure that associations between markers and tested traits are not spurious. The most important source of false associations is the variance distortion attributed to unrecognized population stratification and cryptic relatedness (Cardon and Palmer, 2003). The genomic inflation factor (λ) is the most known measure to determine the success of accounting for population structure. The value of one would reflect the assumption that only a small proportion of loci show a true association (Devlin and Roeder, 1999). In the current study, even after correction for population stratification, λ was still greater than one (e.g. 1.06 for DMY) and the Q-Q plots showed deviations from expected levels (online Supplementary Figure S1). The slight inflation would probably attribute to the polygenic nature of complex traits, in which a large number of genetic variants with small effect affecting the trait variation and/or the trait locus is in LD with multiple genomic regions (Abdel-Shafy *et al.*, 2014).

With 5% Bonferroni genome-wide threshold (P value < 1.1×10^{-6}), we identified 23 significant and 24 suggestive SNPs associated with at least one of the five tested traits (Table 2, and see online Supplementary Fig. S1 and Table S3). These SNPs are located within 36 QTL and distributed over 20 buffalo chromosomes (BBU) with MAF ranges from 0.01 to 0.43. The 47 SNPs were two for DMY, three for FP, 21 for PP, seven for FY, and 14 for PY. One SNP was overlapped for three traits (DMY, FY, and PY); six SNPs were associated with two traits (one for PP and PY; and five for FY and PY); while the rest of markers were associated with only one trait (Table 2 and online Supplementary Table S3). Along with previously reported loci for buffalo studies, the results of current GWAS generally point to the polygenic nature underlying milk production traits in buffalo. However, some chromosomal regions had more prominent associations with relevant traits in corresponding to statistical significance and biological function of nearest candidate genes, making these QTL and genes more likely to be candidates for causal effects. In this respect, our identified SNPs on BBU 1, 5, 6, 7, 9, 19 and 23 reside in the genomic regions where previously reported QTL for milk production traits have been mapped by GWAS in different buffalo populations (de Camargo *et al.*, 2015; El-Halawany *et al.*, 2017; Mokhber, 2017; da Costa Barros *et al.*, 2018; Herrera *et al.*, 2018; Liu *et al.*, 2018). Albeit the same buffalo array was used in these studies, different SNPs were detected. Interestingly, the significant SNPs AX-85128442 and AX-85047828 on BBU 7 (PP: P-values 1.4×10^{-7} and 1.1×10^{-6}, respectively) coincided with previously reported QTL that are repeatedly mapped for milk production traits in Brazilian and Egyptian buffalo (de Camargo *et al.*, 2015; El-Halawany *et al.*, 2017; da Costa Barros *et al.*, 2018). Also, these SNPs were located in genomic regions previously reported for the same traits around corresponding positions in bovine genome (13.2 and 81.0 Mb on BTA6, respectively) in Holstein and Jersey cattle (Cohen-Zinder *et al.*, 2005; Pryce *et al.*, 2010; Buitenhuis *et al.*, 2014; Jiang *et al.*, 2019).

The identified SNPs on BBU 1 (AX-85065545, FY; P-value 9.5×10^{-6}), 5 (AX-85118863 and AX-85095040, FY and PP;

https://doi.org/10.1017/S0022029920000953 Published online by Cambridge University Press

Table 1. Summary statistics of genotyped SNPs in Egyptian buffalo using Axiom Buffalo Genotyping Array

	Number of SNPs			Gap size [kb][a]			LD between SNPs		
Chr	Initial	Filtered[b]	Chr length [Mb]	Mean	SD	Max	Mean r^2	Mean dist [kb]	Max dist [kb]
1	8533	5218	201.95	38.71	23.68	490.26	0.42	145.11	998.03
2	7786	4790	188.83	39.43	28.77	699.06	0.42	154.92	988.67
3	7192	4395	175.39	39.92	27.67	499.56	0.44	152.47	974.82
4	6499	4028	165.14	41.01	32.38	818.93	0.43	158.77	994.68
5	5027	3068	127.53	41.58	41.41	1562.20	0.43	152.31	985.38
6	5064	3103	120.24	38.76	26.37	477.34	0.43	136.62	996.31
7	4722	2878	117.10	40.70	25.01	279.36	0.42	150.18	703.33
8	5031	3061	119.50	39.05	24.45	512.54	0.42	141.70	685.61
9	4341	2698	110.06	40.81	31.29	689.09	0.44	147.65	956.42
10	3929	2447	104.07	42.55	27.40	414.49	0.40	144.73	690.75
11	4012	2460	102.17	41.55	30.54	811.90	0.42	144.64	889.28
12	4461	2763	106.37	38.51	20.64	1,94.63	0.42	136.48	707.51
13	3385	2040	90.22	44.25	51.74	1326.97	0.42	155.84	758.62
14	3342	2058	82.89	40.29	25.60	355.10	0.42	149.38	736.37
15	3428	2109	82.03	38.92	23.16	226.03	0.41	145.37	571.66
16	3185	1937	84.36	43.57	44.47	945.89	0.41	156.53	918.05
17	2937	1755	72.68	41.44	26.45	454.72	0.42	152.44	576.04
18	2737	1662	65.69	39.55	34.47	591.33	0.41	141.91	975.37
19	2968	1839	71.54	38.93	21.27	267.57	0.43	142.39	524.32
20	2637	1638	68.45	41.81	30.34	416.06	0.42	149.88	718.19
21	2557	1628	60.58	37.23	20.14	231.64	0.42	134.61	593.84
22	2529	1563	61.69	39.50	23.44	267.30	0.40	145.03	575.86
23	2172	1332	51.10	38.39	23.91	288.89	0.40	135.53	631.25
24	1779	1109	42.00	37.91	24.26	264.13	0.41	118.80	556.60
25	4891	2590	143.27	55.34	65.82	1397.89	0.51	225.63	998.66
UnPos	17 896	–	–	–	–	–	–	–	–
Overall	123 040	64 169	2614.86	40.77	31.99	1562.20	0.43	151.15	998.66

Chr, chromosome; Mb, mega base; kb, kilo base; LD, linkage disequilibrium; SD, standard deviation; Max, maximum; Min, minimum; r^2, square of the correlation coefficient between pairs of SNPs; dist, distance between the two SNPs that are in LD, UnPos, un-positioned SNPs.
[a]Distance between adjacent SNPs.
[b]Number of SNPs after applying filtering options (see materials and methods for further details) and duplicate SNP position.

P-values 6.9×10^{-6} and 1.2×10^{-8}, respectively) and 6 (AX-85112177, PP; *P*-value 1.6×10^{-9}) were also supported by known QTL in Italian, Brazilian, and Philippine buffalo populations (de Camargo *et al.*, 2015; da Costa Barros *et al.*, 2018; Herrera *et al.*, 2018; Liu *et al.*, 2018). Likewise, they coincided with corresponding locations near to QTL previously mapped in bovine genome in different Holstein and Jersey cattle populations (Rodriguez-Zas *et al.*, 2002; Daetwyler *et al.*, 2008; Pryce *et al.*, 2010; Cole *et al.*, 2011; Buitenhuis *et al.*, 2014).

The detected markers on BBU 9 (AX-85108874 and AX-85120516, PP and PY; *P*-values, 8.5×10^{-7} and 2.6×10^{-6}, respectively), 19 (AX-85103359, PY; *P*-value 9.6×10^{-6}) and 23 (AX-85053529, PP; *P*-value 2.4×10^{-7}) were located in QTL previously discovered in Brazilian, and Iranian buffalo (Mokhber, 2017; da Costa Barros *et al.*, 2018). Also, these SNPs are placed close to QTL reported for several milk production traits in Holstein, Jersey, Brown Swiss, Nordic Red, and Blonde d'Aquitaine cattle breeds (Pryce *et al.*, 2010; Cole *et al.*, 2011; Cecchinato *et al.*, 2012, 2014; Meredith *et al.*, 2012; Iso-Touru *et al.*, 2016; Michenet *et al.*, 2016). Mapping the same genomic regions in independent buffalo and cattle populations with different experimental and analytical structures would increase the confidence that these QTL contain true causative mutations affecting milk production traits.

Beside significant SNPs that overlapped with previously reported loci, several new markers were identified which were not detected in previous buffalo GWAS. Most of these markers are placed on BBU 15, 2, and 4 (Table 2 and online Supplementary Table S3). The most interesting new genomic region is located on BBU 15 between 34.53 and 49.37 Mb,

https://doi.org/10.1017/S0022029920000953 Published online by Cambridge University Press

Table 2. Significant SNPs and candidate genes associated with milk production traits in Egyptian buffalo

Traits	SNP ID	Chr.	Positions [bp]	MA	MAF	β	*P*-value	Nearest gene	Distance [bp]	Other genes	Distance [bp]
DMY	AX-85055593	15	38 842 902	T	0.02	1.417	6.81×10^{-7}	TPD52	114 860	ZBTB10	189 227
FY	AX-85047648	17	27 161 010	A	0.01	0.104	7.99×10^{-7}	ADGRD1	−197 491	SFSWAP	464 689
PP	AX-85112177	6	25 067 300	A	0.15	−0.156	1.58×10^{-9}	SPAG17	Intron		
PP	AX-85070487	13	40 243 654	A	0.06	−0.142	8.33×10^{-9}	TBC1D4	Intron		
PP	AX-85095040	5	105 950 479	C	0.15	0.130	1.25×10^{-8}	KIRREL3	Intron		
PP	AX-85119862	7	72 055 368	T	0.03	−0.379	1.66×10^{-8}	SEL1L3	100 241	RBPJ	196 756
PP	AX-85089717	4	91 165 729	A	0.02	−0.220	1.99×10^{-8}	ATF1	Intron		
PP	AX-85111861	5	59 875 775	T	0.03	0.211	2.12×10^{-8}	SPATA17	−86 737	RRP15	279 437
PP	AX-85044014	15	49 367 898	G	0.03	−0.140	9.33×10^{-8}	C15H8orf34	−223 498	SULF1	313 885
PP	AX-85079904	16	8 665 991	T	0.04	−0.220	1.18×10^{-7}	PHF21A	Intron		
PP	AX-85128442	7	87 744 428	G	0.04	−0.168	1.45×10^{-7}	CAMK2D	−18 204	ANK2	−95 453
PP	AX-85049039	25	50 585 175	G	0.06	−0.156	1.56×10^{-7}	WNK3	Intron		
PP	AX-85109272	15	34 528 577	T	0.01	−0.238	1.93×10^{-7}	EIF3H	−43 945	LOC 102406674	−735 658
PP	AX-85053529	23	20 723 519	G	0.04	−0.268	2.41×10^{-7}	DNMBP	Intron		
PP	AX-85052122	9	48 666 831	G	0.05	0.125	3.20×10^{-7}	LOC102408444	Intron		
PP	AX-85085524	12	58 372 005	C	0.01	−0.260	5.02×10^{-7}	LRRTM4	Intron		
PP	AX-85078085	2	81 255 279	G	0.06	−0.152	7.87×10^{-7}	B3GALT1	Intron		
PP	AX-85108874	9	7 698 653	A	0.01	−0.350	8.49×10^{-7}	NUDT12	370271	LOC 102414284	−709 171
PP	AX-85047828	7	38 018 476	C	0.07	0.130	1.11×10^{-6}	TECRL	−415 231	ADGRL3	−1 737 895
PY	AX-85047648	17	27 161 010	A	0.01	0.049	1.65×10^{-8}	ADGRD1	−197 491	SFSWAP	464 689
PY	AX-85055593	15	38 842 902	T	0.02	0.073	1.72×10^{-7}	TPD52	114 860	ZBTB10	189 227
PY	AX-85111822	15	2 960 480	G	0.01	0.096	6.03×10^{-7}	RALYL	771 376	SNX16	1 489 837
PY	AX-85118863	5	60 745 014	G	0.02	0.096	6.03×10^{-7}	ESRRG	Intron		

Chr, chromosome; MA, minor allele; MAF, minor allele frequency; β, change per minor allele; DMY, daily milk yield; DFY, daily fat yield; DPY, daily protein yield; DLY, daily lactose yield.
In the distance: '+' for upstream and '-' for downstream. Positions are given according to the latest reference assembly of buffalo genome (UOA_WB_1: GCA_003121395.1).

where several SNPs are associated with four milk production traits. One of these SNPs (AX-85055593) is associated with DMY, PY and FY (*P*-values 6.8×10^{-7}, 6.7×10^{-6}, and 1.7×10^{-7}, respectively). Corresponding to bovine genome, this genomic region coincides with previously reported QTL for milk yield as well as protein and fat percentages in Irish and US Holstein along with German Holstein and Charolais crossing cattle (Meredith *et al.*, 2012; Friedrich *et al.*, 2016; Jiang *et al.*, 2019). Interestingly, minor allele of this SNP had a positive effect direction for three traits (1.42, 0.07, and 0.15 units, respectively). On the other hand, our study did not detect most of associations in QTL that have been previously suggested by GWAS in Egyptian buffalo (El-Halawany *et al.*, 2017). This result was probably due to different LD structure of genotyped animals (r^2: 0.43 ± 0.21 *v.* 0.31 ± 0.08), frequency of minor alleles (MAF: 0.29 ± 0.13 *v.* 0.30 ± 0.13), average distance between markers (40.8 ± 32.0 kb *v.* 62.66 ± 67.10), nature of the analyzed trait, and/or potential false associations.

Candidate genes

The location of all identified SNPs in relation to genes on the buffalo genome was calculated according to UOA_WB_1 (GCA_003121395.1) assembly. We found that almost half of SNPs were intronic (51%) while the rest were outside the known genes (Table 2 and online Supplementary Table S3). The nearest genes for those intergenic SNPs are located in a distance ranging from 0.02 to 0.77 Mb. Several identified SNPs are located within or close to numerous candidate genes. For example: the SNP AX-85055593 on BBU 15 associated with DMY, FY and PY is placed 115 and 189 kb far from the genes encoding tumor protein D52 (*TPD52*) and zinc finger and BTB domain containing 10 (*ZBTB10*), respectively. These proteins have functional roles during immune response (Tiacci *et al.*, 2005; Szyda *et al.*, 2019). In addition, the TPD52 is promoting intracellular lipid storage within cultured cells (Kamili *et al.*, 2015). Likewise, the position of the SNP AX-85047648 on BBU 17 that was associated with FY and PY is 197 kb away from the adhesion G protein-coupled receptor D1 (*ADGRD1*) gene. The protein product of this gene is affecting fatty acids concentration in chicken meat (Yang *et al.*, 2018). The SNP AX-85118863 on BBU 5 identified for FY and PY is located in the estrogen related receptor gamma gene (*ESRRG*), which plays a central function in lipid metabolism (Sanoudou *et al.*, 2010). In addition, the SNP AX-85111822 (identified for FY and PY) on BBU 15 is placed at a distance of 0.77 and 1.5 Mb from RNA-binding raly-like (*RALYL*) and sorting nexin 16 (*SNX16*), respectively. These genes have been previously identified as candidates for milk yield and fatty acids content along with resistance to *F. hepatica* (Li *et al.*, 2014; Yodklaew *et al.*, 2017; Twomey *et al.*, 2019). The SNP AX-85043902 on BBU 4 detected for PP and PY is in a distance of 0.55 Mb interval from the gene encoding glutamate receptor interacting protein 1 (*GRIP1*), which is involved in several pathways including metabolic hormones (Hong *et al.*, 1997).

In conclusion, this is the first reported GWAS for milk components in Egyptian buffalo. In the current study, we detected 47 SNPs for five milk production traits (DMY, FP, PP, FY, and PY). These SNPs are located within 36 QTL and distributed over 20 buffalo chromosomes. Some of these loci (11 out of 36) overlap with previously reported QTL in buffalo and/or cattle populations, and some of them are placed within or close to potential candidate genes. The consistence of our identified genomic regions with known QTL and candidate genes provides further evidence for the importance of such loci for the variation in milk production traits. In addition, novel genomic loci were suggested. Further confirmation studies including larger population size should be performed to validate the findings and detect the causal genetic variants.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0022029920000953

Acknowledgments. This study was financially supported by Cairo University, Egypt through the project 'Genomic evaluation for milk production traits in Egyptian buffalo: A step forward for sustainable improvement and food security'.

Financial support. None.

References

Abdel-Shafy H, Awad MAA, El-Regalaty H, Ismael A, El-Assal SE-D and Abou-Bakr S (2020) A single-step genomic evaluation for milk production in Egyptian buffalo. *Livestock Science* **234**, 103977.

Abdel-Shafy H, Bortfeldt RH, Tetens J and Brockmann GA (2014) Single nucleotide polymorphism and haplotype effects associated with somatic cell score in German Holstein cattle. *Genetics Selection Evolution* **46**, 35.

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP and Zondervan KT (2010) Data quality control in genetic case-control association studies. *Nature Protocols* **5**, 1564–1573.

Ardlie KG, Kruglyak L and Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nature Review Genetics* **3**, 299–309.

Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal SE-D and Abdel-Shafy H (2020) Determination of potential candidate genes associated with milk lactose in Egyptian buffalo. *World's Veterinary Journal* **10**, 35–42.

Barłowska J, Szwajkowska M, Litwińczuk Z and Król J (2011) Nutritional value and technological suitability of milk from various animal species used for dairy production. *Comprehensive Review in Food Science and Food Safety* **10**, 291–302.

Brodie A, Azaria JR and Ofran Y (2016) How far from the SNP may the causative genes be? *Nucleic Acids Research* **44**, 6046–6054.

Buitenhuis B, Janss LL, Poulsen NA, Larsen LB, Larsen MK and Sorensen P (2014) Genome-wide association and biological pathway analysis for milkfat composition in Danish Holstein and Danish Jersey cattle. *BMC Genomics* **15**, 1112.

Cardon LR and Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet (London, England)* **361**, 598–604.

Cecchinato A, Ribeca C, Maurmayr A, Penasa M, De Marchi M, Macciotta NP, Mele M, Secchiari P, Pagnacco G and Bittante G (2012) Short communication: effects of beta-lactoglobulin, stearoyl-coenzyme A desaturase 1, and sterol regulatory element binding protein gene allelic variants on milk production, composition, acidity, and coagulation properties of Brown Swiss cows. *Journal of Dairy Science* **95**, 450–454.

Cecchinato A, Ribeca C, Chessa S, Cipolat-Gotet C, Maretto F, Casellas J and Bittante G (2014) Candidate gene association analysis for milk yield, composition, urea nitrogen and somatic cell scores in Brown Swiss cows. *Animal: An International Journal of Animal Bioscience* **8**, 1062–1070.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM and Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7. Available at www.cog-genomics.org/plink/1.9/

Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, der Wind A E-V, Lee JH, Drackley JK, Band MR, Hernandez AG, Shani M, Lewin HA, Weller JI and Ron M (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research* **15**, 936–944.

Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ Jr., Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK and Da Y (2011) Genome-wide association analysis of thirty one production, health,

reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* **12**, 408

da Costa Barros C, de Abreu Santos DJ, Aspilcueta-Borquis RR, de Camargo GMF, de Araujo Neto FR 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.

Daetwyler HD, Schenkel FS, Sargolzaei M and Robinson JA (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *Journal of Dairy Science* **91**, 3225–3236.

de Camargo GM, Aspilcueta-Borquis RR, Fortes MR, Porto-Neto R, Cardoso DF, Santos DJ, Lehnert SA, Reverter A, Moore SS and Tonhati H (2015) Prospecting major genes in dairy buffaloes. *BMC Genomics* **16**, 872.

Deb GK, Nahar TN, Duran PG and Presicce GA (2016) Safe and sustainable traditional production: the water buffalo in Asia. *Frontiers in Environmental Science* **4**, 38.

Devlin B and Roeder K (1999) Genomic control for association studies. *Biometrics* **55**, 997–1004.

El-Halawany N, Abdel-Shafy H, Shawky AA, Abdel-Latif MA, Al-Tohamy AFM and Abd El-Moneim OM (2017) Genome-wide association study for milk production in Egyptian buffalo. *Livestock Science* **198**, 10–16.

FAO (2019) FAOSTAT database collections. Food and Agriculture Organization of the United Nations. Available at http://www.fao.org/faostat/

Friedrich J, Brand B, Ponsuksili S, Graunke KL, Langbein J, Knaust J, Kuhn C and Schwerin M (2016) Detection of genetic variants affecting cattle behaviour and their impact on milk production: a genome-wide association study. *Animal Genetics* **47**, 12–18.

Goddard ME and Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Review Genetics* **10**, 381–391.

Hayes BJ and Daetwyler HD (2019) 1000 Bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. *Annual Review of Animal Bioscience* **7**, 89–102.

Herrera JRV, Flores EB, Duijvesteijn N, Gondro C and van der Werf JHJ (2018) Genome-wide association study for milk traits in Philippine dairy buffaloes. In *Proceedings of the 11th World Congress on Genetics Applied to Livestock Production, 11–16 February 2019, Auckland, New Zealand*, p. 825.

Hong H, Kohli K, Garabedian MJ and Stallcup MR (1997) GRIP1, a transcriptional coactivator for the AF-2 transactivation domain of steroid, thyroid, retinoid, and vitamin D receptors. *Molecular and Cellular Biology* **17**, 2735–2744.

Hu ZL, Park CA and Reecy JM (2019) Building a livestock genetic and genomic information knowledgebase through integrative developments of animal QTLdb and CorrDB. *Nucleic Acids Research* **47**, D701–D710.

Iamartino D, Nicolazzi EL, Van Tassell CP, Reecy JM, Fritz-Waters ER, Koltes JE, Biffani S, Sonstegard TS, Schroeder SG, Ajmone-Marsan P, Negrini R, Pasquariello R, Ramelli P, Coletta A, Garcia JF, Ali A, Ramunno L, Cosenza G, de Oliveira DAA, Drummond MG, Bastianetto E, Davassi A, Pirani A, Brew F and Williams JL (2017) Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). *PLoS ONE* **12**, e0185220.

Iso-Touru T, Sahana G, Guldbrandtsen B, Lund MS and Vilkki J (2016) Genome-wide association analysis of milk yield traits in Nordic red cattle using imputed whole genome sequence variants. *BMC Genetics* **17**, 55.

Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB and Da Y (2019) A large-scale genome-wide association study in U.S. Holstein cattle. *Frontiers in Genetics* **10**, 412.

Kamili A, Roslan N, Frost S, Cantrill LC, Wang D, Della-Franca A, Bright RK, Groblewski GE, Straub BK, Hoy AJ, Chen Y and Byrne JA (2015) TPD52 expression increases neutral lipid storage within cultured cells. *Journal of Cell Science* **128**, 3223–3238.

Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, Boehm F, Caporaso NE, Cornelis MC, Edenberg HJ, Gabriel SB, Harris EL, Hu FB, Jacobs KB, Kraft P, Landi MT, Lumley T, Manolio TA, McHugh C, Painter I, Paschall J, Rice JP, Rice KM, Zheng X and Weir BS (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology* **34**, 591–602.

Li C, Sun D, Zhang S, Wang S, Wu X, Zhang Q, Liu L, Li Y and Qiao L (2014) Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. *PLoS ONE* **9**, e96186.

Liu JJ, Liang AX, Campanile G, Plastow G, Zhang C, Wang Z, Salzano A, Gasparrini B, Cassandro M and Yang LG (2018) Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. *Journal of Dairy Science* **101**, 433–444.

Lu XR, Duan AQ, Li WQ, Abdel-Shafy H, Rushdi HE, Liang SS, Ma XY, Liang XW and Deng TX (2020) Genome-wide analysis reveals genetic diversity, linkage disequilibrium, and selection for milk production traits in Chinese buffalo breeds. *Journal of Dairy Science* **103**, 4545–4556.

Meredith BK, Kearney FJ, Finlay EK, Bradley DG, Fahey AG, Berry DP and Lynn DJ (2012) Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genetics* **13**, 21.

Michenet A, Barbat M, Saintilan R, Venot E and Phocas F (2016) Detection of quantitative trait loci for maternal traits using high-density genotypes of Blonde d'Aquitaine beef cattle. *BMC Genetics* **17**, 88.

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T and Lee DH (2002) BLUPF90 and related programs (BGF90). In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, France*, p. 28.

Mokhber M (2017) Genome-wide association study for milk production in Iranian buffalo. In *Proceedings of the 1st International and 5th National Conference on Organic vs. Conventional Agriculture, 16–17 August 2017, Ardabil, Iran*, pp. 1–5.

Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME and Hayes BJ (2010) A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *Journal of Dairy Science* **93**, 3331–3345.

Rodriguez-Zas SL, Southey BR, Heyen DW and Lewin HA (2002) Detection of quantitative trait loci influencing dairy traits using a model for longitudinal data. *Journal of Dairy Science* **85**, 2681–2691.

SADS (2009) Sustainable Agricultural Development Strategy towards 2030. Agricultural Research and Development Council, Ministry of Agriculture and Land Reclamation, Arab Republic of Egypt, p. 194.

Sanoudou D, Duka A, Drosatos K, Hayes KC and Zannis VI (2010) Role of Esrrg in the fibrate-mediated regulation of lipid metabolism genes in human ApoA-I transgenic mice. *Pharmacogenomics Journal* **10**, 165–179.

Szyda J, Mielczarek M, Fraszczak M, Minozzi G, Williams JL and Wojdak-Maksymiec K (2019) The genetic background of clinical mastitis in Holstein-Friesian cattle. *Animal: An International Journal of Animal Bioscience* **13**, 2156–2163.

Tiacci E, Orvietani PL, Bigerna B, Pucciarini A, Corthals GL, Pettirossi V, Martelli MP, Liso A, Benedetti R, Pacini R, Bolli N, Pileri S, Pulford K, Gambacorta M, Carbone A, Pasquarello C, Scherl A, Robertson H, Sciurpi MT, Alunni-Bistocchi G, Binaglia L, Byrne JA and Falini B (2005) Tumor protein D52 (TPD52): a novel B-cell/plasma-cell molecule with unique expression pattern and Ca(2+)-dependent association with annexin VI. *Blood* **105**, 2812–2820.

Turner SD (2014) qqman: an R package for visualizing GWAS results using QQ and Manhattan plots. https://doi.org/10.1101/005165

Twomey AJ, Berry DP, Evans RD, Doherty ML, Graham DA and Purfield DC (2019) Genome-wide association study of endo-parasite phenotypes using imputed whole-genome sequence data in dairy and beef cattle. *Genetics Selection Evolution* **51**, 15.

VanRaden PM and Wiggans GR (1991) Derivation, calculation, and use of national animal model information. *Journal of Dairy Science* **74**, 2737–2746.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA and Yang J (2017) 10 Years of GWAS discovery: biology, function, and translation. *American Journal of Human Genetics* **101**, 5–22.

Wanapat M, Ngarmsang A, Korkhuntot S, Nontaso N, Wachirapakorn C, Beakes C and Rowlinson P (2000) A comparative study on the rumen microbial population of cattle and swamp buffalo raised under traditional village conditions in the Northeast of Thailand. *Asian-Australasian Journal of Animal Science* **13**, 918–921.

Yang S, Wang Y, Wang L, Shi Z, Ou X, Wu D, Zhang X, Hu H, Yuan J, Wang W, Cao F and Liu G (2018) RNA-Seq reveals differentially expressed genes affecting polyunsaturated fatty acids percentage in the Huangshan Black chicken population. *PLoS ONE* **13**, e0195132.

Yodklaew P, Koonawootrittriron S, Elzo MA, Suwanasopee T and Laodim T (2017) Genome-wide association study for lactation characteristics, milk yield and age at first calving in a Thai multibreed dairy cattle population. *Agriculture Natural Resources* **51**, 223–230.

https://doi.org/10.1017/S0022029920000953 Published online by Cambridge University Press