2020, Scienceline Publication
World's Veterinary Journal
World Vet J, 10(1): 35-42, March 25, 2020
ISSN 2322-4568
DOI: https://dx.doi.org/10.36380/scil.2020.wvj5

Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo

Mohamed A.A. Awad[1], Samy Abou-Bakr[1], Hussein El-Regalaty[2], Salah El-Din El-Assal[3] and Hamdy Abdel-Shafy[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.*
[3]*Department of Genetics, Faculty of Agriculture, Cairo University, El-Gamma street, 12613 Giza, Egypt.*

*Corresponding author's Email: hamdyabdelshafy@agr.cu.edu.eg; ORCID: 0000-0003-3282-3899

ORGINAL ARTICLE
pii: S232245682000005-10
Received: 17 Jan 2020
Accepted: 19 Feb 2020

ABSTRACT

The aim of the present genome-wide association study (GWAS) was to identify single nucleotide polymorphisms (SNPs) and candidate genes associated with lactose percentage (LP) and lactose yield (LY) in Egyptian buffalo. The phenotypic dataset included 60,318 monthly measures for LP and LY from 1481 animals. A total number of 114 animals with high and low deviated performance were selected for genotyping with Axiom Buffalo Genotyping 90K Array. Genome-wide analysis was performed using a single marker regression. The GWAS revealed 32 significant and seven suggestive SNPs for LP, however; only two suggestive SNPs were identified for LY. The identified genomic regions are overlapped with previously reported QTL in different cattle breeds. In addition, novel genomic loci were detected. The identified genomic regions harbored many candidate genes with biological roles associated with milk production traits, such as TPD52 and ZBTB10 on chromosome 15; AADAT and GALNTL6 on chromosome 3 and COL8A1 and PLOD2 on chromosome 1. Our findings provide the basis to uncover the key markers and candidate genes affecting lactose traits which facilitate the exploration of the genetic mechanisms that control lactose traits variation in Egyptian buffalo.

Key words: Candidate gene, Egyptian buffalo, Genome, Genomic loci, Lactose

INTRODUCTION

Buffalo milk contains higher level of total solids (fat, proteins, lactose and ash) compared to other farm animal species, with percentage of 12-25.5% (Gantner et al., 2015). The total solids of buffalo milk make it ideal for processing into dairy products due to its high quality. Lactose is the second major constituent of buffalo milk with percentage of 3.2-4.9% (Ménard et al., 2010). According to previous studies, lactose has a great impact on milk production due to its osmotic characteristic that helps to pull the water into the mammary epithelial cells (Lin et al., 2016). In dairy cattle, lactose concentration was found to be a good indicator for energy balance (Reist et al., 2002), pregnancy rate (Buckley et al., 2003), udder health (Ptak and Bieniek, 2012), immunity, and longevity (Miglior et al., 2006).

Milk production is a quantitative and complex trait; *i.e.* controlled by a large number of genes with small effects and affected by several environmental factors (Hill, 2012). This complexity makes it difficult to understand the biological and the genetic mechanisms that control the trait variation by using traditional breeding (Hill, 2012).

Recently, with the development of molecular biotechnology including sequencing of whole genome for many livestock species, it becomes possible to obtain genomic information including thousands of single nucleotide polymorphisms (SNPs) covering the whole genome. This is followed by rapid development of SNPs genotyping chips (Iso-Touru et al., 2016). The availability of SNP genotyping chips makes it possible to perform genome-wide association studies (GWAS) to identify significant genomic regions associated with the trait of interest. Therefore, GWAS increases the power to map quantitative trait loci (QTL) and defines narrower genomic regions that harbor causal genes associated with economically important traits (Bouwman et al., 2011).

In cattle, several GWAS were conducted to identify SNPs related to many of economically important traits such as body conformation (Wu et al., 2013), disease resistance (Finlay et al., 2012), growth traits (Bolormaa et al., 2011), fertility traits (Huang et al., 2010) and milk production traits (Nayeri et al., 2016). In buffalo, the commercial cattle SNP chips were used to explore genomic regions that are associated with milk production traits, since buffalo and cattle are closely related (Venturini et al., 2014). More recently, the availability of buffalo's reference genome opens the field for developing and releasing a commercial buffalo SNP chip (Axiom® Buffalo Genotyping 90K Array).

To cite this paper: Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal S.E-D and Abdel-Shafy H (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo. *World Vet. J.*, 10 (1): 35-42. DOI: https://dx.doi.org/10.36380/scil.2020.wvj5

To our knowledge, few GWAS have been performed so far using this chip. Two GWAS were conducted in each Italian (Iamartino et al., 2013 and Liu et al., 2018), Brazilian (de Camargo et al., 2015 and Gonzalez Guzman et al., 2020), and Egyptian buffalo (El-Halawany et al., 2017 and Abdel-Shafy et al., 2020); while one study was performed in each Philippine (Herrera et al., 2018) and Iranian buffalo (Mokhber et al., 2019). The results of such studies showed the benefits of GWAS to identify genomic regions associated with milk production traits and facilitate the utilization of genetic potential for improvement of Egyptian buffalo milk performance.

Therefore, the objective of this study was to perform a genome-wide association study using Axiom Buffalo Genotyping 90K Array to identify SNP markers and potential candidate genes associated with lactose percentage and lactose yield in Egyptian buffalo.

MATERIALS AND METHODS

Animals and phenotypes

Milk samples from 1481 animals (50 ml each) were monthly collected. The samples were maintained frozen at -20◦ C until performed the chemical analysis of milk constituents. Percentage of lactose was determined by Infrared Milk Analyzer (Bentley, I50®) and the automated method of infrared absorption spectrophotometry (Milk-o-Scan; Foss Electric, Hillerφd, Denmark) at Cattle Information System/Egypt (CISE) and Animal Production Research Institute (APRI) of Agricultural Research Center (ARC). Records for chemical composition of milk were checked carefully to remove abnormal phenotypic values and exclude animals that have less than three times chemical analyses per parity. After quality check, a total number of 60,318 monthly measures were remained. The average lactose percentage was 5.1±0.6. Yield of lactose was calculated by multiplying lactose percentage by milk yield at the same day. The average lactose yield was 0.44±0.15 kg/day.

Ethical approval

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

Genotypic data

From the 1481 individuals, we selected the highest and lowest 114 animals for genotyping according to average daily milk yield. Blood samples were collected from the jugular vein of these animals and kept in a 15 ml Falcon tube containing 1 ml 0.5 M EDTA as an anticoagulant. The samples were immediately placed on the cooling gel in an ice box after their collection and transferred to the laboratory and kept away from the direct sunlight. The samples temporarily were stored at -20°C before DNA extraction. Genomic DNA was extracted from whole blood samples using a QIAamp® DNA Blood Mini Kit (QIAGEN, Hilden, Germany).

Using Axiom® Buffalo Genotyping 90K array, genotyping was performed according to the standard protocol of Thermo Fisher Scientific. The signal intensity from the raw genotypic data (CEL files) were converted into genotype calls and annotated to the reference assembly of buffalo genome using Genotyping Console™ 4.2. The quality of the raw genotypes was checked using PLINK 1.9 (Chang et al., 2015). In this respect, the SNPs with unknown positions on buffalo genome were eliminated. In addition, SNPs with low call rate (missing genotype per SNP >0.15), with low minor frequency (MAF<0.01), and/or deviated markers from Hardy-Weinberg proportion (P<0.0001) were excluded. Individuals with low call rate (P<0.15) were also discarded. After applying 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%. Some pairs of SNPs may have complete linkage disequilibrium (LD), and thus would convey similar information (Anderson et al., 2010 and Laurie et al., 2010). Therefore, we excluded one of a pair of these 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 pruning step led to reduce the SNP number to 44,985 markers.

Statistical analysis

a. Initial analysis

To determine the significant factors affecting the traits, we initially tested all available fixed effects using general linear model procedure in R program as follow: $y_{ijklmno} = \mu + MF_i + Lac_j + H_k + YS_l + AC_m + DIM_n + \varepsilon_{ijklmno}$ where, $y_{ijklmno}$ is the phenotypic observations (lactose percentage and yield); μ is the overall mean of observations, MF_i is the effect of i^{th} milking/day (i=1 to 3); Lac_j is the effect of j^{th} parity number (j=1 to 13); H_k is the effect of k^{th} herd (11 herds); YS_l is the combined effect of l^{th} year and season of calving (68 levels); AC_m is the effect of m^{th} age at calving (1,341 levels); DIM_n is the effect of n^{th} days in milk in each parity (285 levels); and $\varepsilon_{igklmno}$ is the residual error. Since all tested variables showed significant effects, we used all for the next animal model to calculate the yield deviations.

To cite this paper: Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal S.E-D and Abdel-Shafy H (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo. *World Vet. J.*, 10 (1): 35-42. DOI: https://dx.doi.org/10.36380/scil.2020.wvj5

b. Phenotype adjustment

A yield deviation is defined as a weighted average of animal's own performance adjusted for non-genetic factors. This procedure is used to initially adjust phenotypes before GWAS to reduce the residual error (VanRaden and Wiggans, 1991). In the current investigation, yield deviations are estimated for each trait as the sum of breeding value and residual for each animal. Estimated breeding values and residuals for each trait were computed by univariate animal model using BLUPF90 family (Misztal et al., 2002). Given the vector of y representing the phenotypic observations on the tested trait (lactose percentage and yield), the following univariate animal model was used: $y = X_b + Z_\alpha + W_p + \varepsilon$, where b is the vector of all fixed effects including milking frequency per day, parity number, herd, year and season of calving. Linear regressions of age at calving and the fourth order Legendre polynomials of DIM were also 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 connecting observations of y to fixed, random animal, and random permanent environmental effects, respectively.

Genome-wide association analyses

A potential problem associated with population structure was adjusted with a multidimensional scaling (MDS) approach implemented in PLINK 1.9 (Chang et al., 2015). In this regards, the scaling process led to eight significant clusters representing axes of ancestry at P<0.0001. These clusters were used as covariates in the model when performing GWAS. The GWAS was performed using the linear regression model in PLINK 1.9 (Chang et al., 2015), where the adjusted lactose percentage and yield were regressed on the number of copies of the alleles using PLINK–linear option with population stratification as covariates. The results of associations were used to generate Manhattan and Q-Q plots using SNPEVG (Wang et al., 2012). To prevent false positive signals, Bonferroni correction was applied to adjust for multiple testing.

QTL and candidate genes

Since the buffalo and cattle are closely related, previously reported QTL for lactose percentage and yield were retrieved from animal QTLdb (http://www.animalgenome.org/QTLdb), release 37 (Hu et al., 2019). Candidate genes in each genomic region were extracted from the latest annotated file (na35.r2.a2) of Axiom® Buffalo Genotyping array provided by Thermo Fisher Scientific (2019).

RESULTS AND DISCUSSION

Genotypes for genome-wide scan

The Axiom Buffalo Genotyping Array used in this study featured 123,040 SNPs spanning the entire buffalo genome with an average spacing of one SNP every 34.46 kb across all loci (median spacing of 29.54 kb, a minimum distance of 0.01 kb and a maximum distance of 1.33 Mb). After quality control procedures, a total of 44,985 SNPs (36.6%) and 113 animals were remained for further analysis. This subset of SNPs covered 2,614.86 Mb of the buffalo genome with the shortest length of 42.13 Mb for chromosome 24 and longest length of 201.95 Mb for chromosome 1. The average physical distance between markers was 40.77 kb (median spacing of 31.60 kb, a minimum distance of 0.01 kb and a maximum distance of 1.56 Mb). The distribution of SNPs varied among the chromosomes, where the number of SNPs per chromosome were ranged from 1109 (chromosome 24) to 5218 (chromosome 1). While, the SNP density were ranged from 18.1 SNP/Mb (on chromosome 25) to 26.9 SNP/Mb (on chromosome 21) with an average density of 24.7 SNP/Mb. The MAF was ranged from 0.01 to 0.5 with an average of 0.29, which indicates the existence of variation in the allele frequency among the SNPs markers in current study.

Assessment of population structure for GWAS

To assess the successful correction for false positive associations resulted from population structure, the genomic inflation factor (λ) and quantile-quantile (Q-Q) plot were used (Power et al., 2016). Under null hypothesis of no association, λ should equal to one. Price et al. (2010) reported that a λ value less than 1.05 is acceptable for association studies. For Q-Q plot, it should exhibit y = x distribution under null hypothesis of no association (Power et al., 2016). In the current study, even after the correction for population structure, the λ was 1.08 and 1.28 for lactose yield (LY) and lactose percentage (LP), respectively; and the Q-Q plots showed a large deviation from the expected 1:1 relationship (Figure 1 and 2). This inflation may be due to many reasons; firstly, when a large number of loci showed a strong association with the trait (Guo et al., 2012); secondly, hidden relatedness between genotyped animals; thirdly, the low heritability and the nature of lactose traits as polygenic quantitative traits affected by many genes, each with small effects (Power et al., 2016); and fourthly, higher LD between evaluated SNPs (Abdel-Shafy et al., 2014).

To cite this paper: Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal S.E-D and Abdel-Shafy H (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo. *World Vet. J.*, 10 (1): 35-42. DOI: https://dx.doi.org/10.36380/scil.2020.wvj5

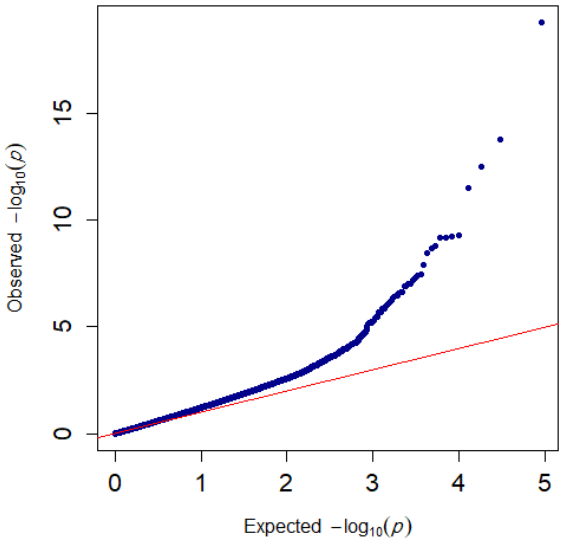

Figure 1. Quantile-Quantile plot of genome-wide association for lactose percentage in Egyptian buffalo

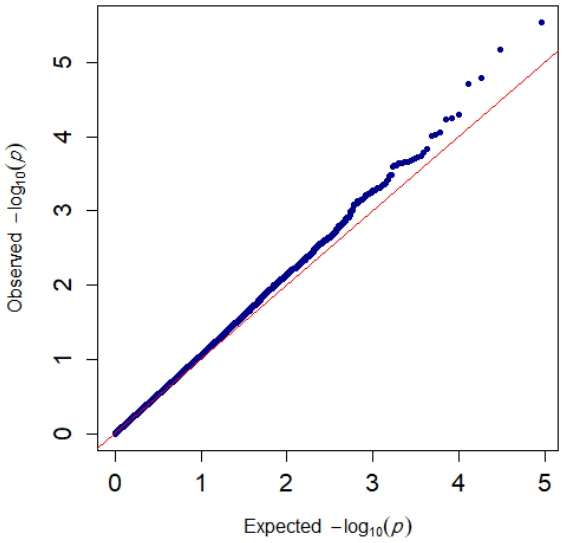

Figure 2. Quantile-Quantile plot of genome-wide association for lactose yield in Egyptian buffalo

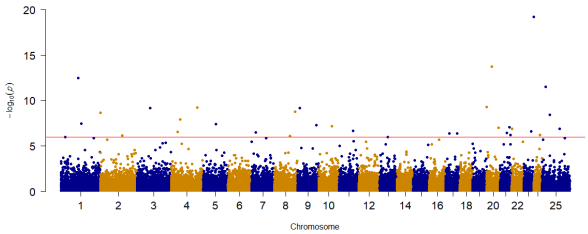

Figure 3. Manhattan plot of genome-wide association for lactose percentage in Egyptian buffalo.
The horizontal red line indicates the whole-genome significance threshold after Bonferroni correction at $\alpha = 0.05$ ($P \leq 1.11 \times 10^{-6}$)

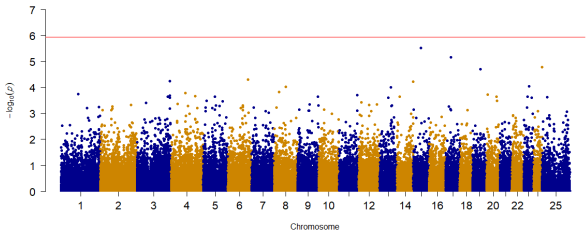

Figure 4. Manhattan plot of genome-wide association for lactose yield in Egyptian buffalo.
The horizontal red line indicates the whole-genome significance threshold after Bonferroni correction at $\alpha = 0.05$ ($P \leq 1.11 \times 10^{-6}$)

Genome-wide association for lactose percentage

Manhattan plot exhibiting P-values of genome wide scan with respect to genomic location for each trait are shown in Figure 3 and 4. Manhattan plot presents the negative logarithm of P-value for all tested SNPs across the genome (y-axis) versus its chromosomal locations (x-axis). The horizontal red line indicates the whole-genome significance threshold after Bonferroni correction at $\alpha = 0.05$ ($P \leq 1.11 \times 10^{-6}$). A strong association with the trait is represented with a small P-value, thus a negative logarithm value will have high value and the scatter plot will be at the highest peak along the y-axis in the Manhattan plot. In the current instigation, the GWAS for LP revealed 32 significant SNPs distributed on 18 chromosomes (Figure 3 and Table 1). These SNPs were occurred with MAF of 0.01 to 0.09. The most significant SNP (AX-85107786) for association with LP was located on chromosome 23 at position 50,923,849 bp with P-value of 5.90×10^{-20}. This SNP was located at 29,570 bp upstream from the ADGRA1 gene and 72,007 bp upstream from the CFAP46 gene with MAF of 0.01 and effect size of -0.429 units of LP. The effect size of minor alleles for all significant SNPs was negative, ranging from -0.861 to -0.195 units of LP, indicating their association with lower LP. In addition, seven suggestive SNPs were identified for association with LP. All these SNPs were occurred with MAF of 0.01 to 0.11. Interestingly, the effect of minor alleles for the top four suggestive SNPs on chromosomes 13, 25, 7 and 1 was negative, ranging from -0.427 to -0.142 units of LP, while the effect of minor alleles for the other three SNPs on chromosomes 2, 16 and 25 was the same and had a positive effect direction for LP (0.438 units).

To cite this paper: Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal S.E-D and Abdel-Shafy H (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo. *World Vet. J.*, 10 (1): 35-42. DOI: https://dx.doi.org/10.36380/scil.2020.wvj5

Table 1. Significant single nucleotide polymorphisms associated with lactose traits and candidate genes in Egyptian buffalo

Traits	SNP ID	Chr.	Positions (bp)	MA	MAF	β	P-value	Nearest gene	Distance (bp)	Other genes	Distance (bp)
LP	AX-85107786	23	50923849	G	0.01	-0.429	5.90E-20	ADGRA1	+29570	CFAP46	+72007
LP	AX-85075099	20	27966184	C	0.02	-0.861	1.79E-14	STRN3	Intron		
LP	AX-85088945	1	88740750	G	0.02	-0.611	3.18E-13	COL8A1	Intron		
LP	AX-85067933	25	19690758	A	0.01	-0.348	3.17E-12	SMPX	+6973	KLHL34	-58203
LP	AX-85102661	20	1344536	A	0.02	-0.615	5.27E-10	CKB	+1183	TRMT61A	+3744
LP	AX-85071608	4	135416124	C	0.01	-0.475	5.82E-10	ANK3	Intron		
LP	AX-85061010	9	11244996	G	0.01	-0.447	6.62E-10	FAM174A	+361585	CHD1	+1286215
LP	AX-85083471	3	66737950	C	0.02	-0.461	6.96E-10	AADAT	+81516	GALNTL6	+1869717
LP	AX-85086185	8	107090040	A	0.03	-0.377	1.77E-09	LOC102404258	Intron		
LP	AX-85053506	2	1531993	A	0.04	-0.279	2.26E-09	GMDS	-43447	MYLK4	+173684
LP	AX-85067145	25	41019921	T	0.06	-0.552	3.78E-09	PPP1R2C	+323714	CASK	-543346
LP	AX-85125967	4	46892040	A	0.02	-0.281	1.22E-08	ISX	+319202	LARGE1	+615445
LP	AX-85044502	1	105132869	T	0.03	-0.297	3.50E-08	GAP43	Intron		
LP	AX-85091884	5	64880712	A	0.02	-0.335	4.02E-08	BRINP3	-972845	KCTD3	+2382794
LP	AX-85093722	9	96947301	G	0.09	-0.195	5.40E-08	LOC102404549	-10377	LOC102404027	-24596
LP	AX-85105519	10	68030846	C	0.04	-0.260	6.46E-08	HS3ST5	Intron		
LP	AX-85109519	21	48673809	C	0.02	-0.467	9.30E-08	WDR82	+7122	GLYCTK	+25254
LP	AX-85098567	20	62110044	A	0.01	-0.433	9.62E-08	TTC23	Intron		
LP	AX-85055462	25	92157984	T	0.01	-0.403	1.23E-07	ZMAT1	+9638	LOC102402428	-30612
LP	AX-85121534	22	851096	A	0.01	-0.239	1.29E-07	BCL2	-43867	PHLPP1	-65403
LP	AX-85079804	11	71648876	G	0.02	-0.407	2.34E-07	DPH6	+527178	C11H15orf41	+704549
LP	AX-85088569	23	37120921	C	0.02	-0.325	2.41E-07	LOC112581570	+9890	PDZD8	-98381
LP	AX-85078270	4	33866814	C	0.04	-0.314	2.76E-07	SOX5	Intron		
LP	AX-85049384	7	22266031	T	0.02	-0.444	3.47E-07	CFAP299	Intron		
LP	AX-85049679	21	34720210	C	0.03	-0.321	3.66E-07	KBTBD8	+215342	LRIG1	+266570
LP	AX-85098463	17	58602702	A	0.06	-0.204	4.05E-07	GAB1	Intron		
LP	AX-85116698	17	19620490	A	0.04	-0.303	4.49E-07	WDR66	Intron		
LP	AX-85053490	21	53297221	C	0.01	-0.433	6.10E-07	CCR1	-613	XCR1	+64141
LP	AX-85086544	24	31016341	T	0.03	-0.319	6.43E-07	CPPED1	Intron		
LP	AX-85070682	2	114150095	T	0.09	-0.432	7.14E-07	DARS	Intron		
LP	AX-85085611	8	81186163	T	0.03	-0.397	8.74E-07	YAE1	+133461	POU6F2	-207100
LP	AX-85093634	1	22545784	A	0.01	-0.429	1.02E-06	LOC102402668	-64615	SGCZ	-102736
LP	AX-85124326*	13	40601425	T	0.11	-0.142	1.12E-06	TBC1D4	-153107	LOC112578532	-1093876
LP	AX-85078356*	25	117341243	C	0.05	-0.258	1.37E-06	CT83	-262287	LOC102398514	+377742
LP	AX-85058363*	7	75945732	G	0.02	-0.427	1.40E-06	KCNIP4	+18762	ADGRA3	-394941
LP	AX-85067191*	1	168416519	T	0.06	-0.195	1.44E-06	C1H3orf58	-980648	PLOD2	-1388647
LP	AX-85052475*	2	37072441	A	0.01	0.438	1.98E-06	CLMP	Intron		
LP	AX-85078483*	16	51213059	T	0.01	0.438	1.98E-06	LOC112578406	+30121	FOXP4	+119159
LP	AX-85070334*	25	5377908	A	0.02	0.438	1.98E-06	LOC102403725	Intron	CFAP46	+72007
LY	AX-85055593*	15	38842902	T	0.02	0.113	2.92E-06	TPD52	+114860	ZBTB10	+189227
LY	AX-85047648*	17	27161010	A	0.01	0.068	6.76E-06	ADGRD1	-197491	SFSWAP	+464689

Chr: chromosome, MA: minor allele, MAF: minor allele frequency, β: change per minor allele, LP: lactose percentage, LY: lactose yield, bp: base pair, SNP: single nucleotide polymorphisms,* Suggestive SNPs (SNPs that are close to significant threshold line based of Bonferroni correction). 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).

To cite this paper: Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal S.E-D and Abdel-Shafy H (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo. *World Vet. J.*, 10 (1): 35-42. DOI: https://dx.doi.org/10.36380/scil.2020.wvj5

The association results for LP not only provide confirmatory evidences for previously findings, but also explore a suite of novel significant SNPs that did not reported by previous association or linkage studies in buffalo and cattle populations. Among the significant SNPs that were supported by previously reported QTL in the cattle QTL database; SNP (AX-85067933) which is located on chromosome 25 at position 19,690,758 bp. This SNP is corresponding to the position of 128,877,968 bp on chromosome x in the bovine genome and close to previously detected SNP at 137.1 Mb for LP in Holstein and Jersey cattle (Benedet et al., 2019). Furthermore, the significant SNP (AX-85071608) on chromosome 4 at position 135,416,124 bp is corresponding to the position of 16,073,463 bp on BTA 28 in bovine genome. This position resides very close to previously reported QTL for LP that has been mapped using GWAS in Holstein and Jersey cattle at 16.3 Mb (Benedet et al., 2019). The three significant SNPs located on chromosome 21 between 34.72 and 53.29 Mb are corresponding to the position between 34.80 and 53.85 Mb on chromosome 22 in bovine genome. These genomic regions reside close to previously reported QTL for LP that have been mapped by linkage study in Chinese Holstein cattle at 38.8-39.0 Mb (Mao et al., 2015).

Genome-wide association for lactose yield

The GWAS for LY did not show any significant associations. However, GWAS identified two suggestive SNPs were close to the Bonferroni corrected threshold for genome wide significance ($\alpha = 0.05$, $P \leq 1.11 \times 10^{-6}$) for association with LY (Figure 4 and Table 1). The highest peak was observed on chromosome 15 for the SNP (AX-85055593) at position 38,842,902 bp (P-value of 2.92×10^{-6}). This SNP was located at 114,860 bp upstream from the TPD52 gene and 189,227 bp upstream from the ZBTB10 gene with MAF of 0.02 and effect size of 0.113 units of LY. On chromosome 17, the second suggestive SNP (AX-85107786) was located at position 50,923,849 bp with P-value of 4.00×10^{-6}, MAF of 0.01, and effect size of 0.068 units of LY. This SNP was located at 19,7491bp downstream from the ADGRD1 gene and 464,689 bp upstream from the SFSWAP gene. The minor alleles of these suggestive SNPs (AX-85055593 and AX-85107786) were favorable and associated with higher LY. These SNPs are considered novel genomic loci for LY since they were not detected in previous buffalo and cattle GWAS or linkage studies.

Identification of possible candidate genes

After conducting the association analyses to identify the top SNPs associated with studied traits and refine their positions, we used these positions to examine nearby genes in order to identify potential candidate genes affecting the relevant traits. Selection of potential candidate genes was based on previously reported QTL and the biological functions related to milk production, milk composition and immune response. For example, the candidate genes COL8A1 and PLOD2 on chromosome 1, CCR1 and LRIG1 on chromosome 21, CHD1 on chromosome 9 and ZBTB10 on chromosome 15, are associated with immune response of mammary gland, somatic cell count and mastitis resistance (Chen et al., 2015; Fang et al., 2016; Banos et al., 2017; Welderufael et al., 2018 and Szyda et al., 2019). Furthermore, the candidate gene MYLK4 on chromosome 2 was previously reported to have a significant effect on milk yield, fat percentage, fat yield and protein yield in Brazilian buffaloes (de Camargo et al., 2015). The identified region on chromosome 4 contains LARGE1 gene that is associated with daily milk yield in Egyptian buffaloes (El-Halawany et al., 2017). In addition, the candidate genes AADAT and GALNTL6 gene on chromosome 3 were previously reported to be associated with myristic saturated fatty acid content in the meat, feed efficiency and growth traits in Ireland Holstein-Friesian and Hereford cattle (Doran et al., 2014 and Seabury et al., 2017).

CONCLUSION

This is the first GWAS for lactose traits in Egyptian buffalo. Our findings provide the basis to uncover genomic regions associated with lactose traits in Egyptian buffalo. These genomic regions coincided with previously reported QTL for milk production traits in different cattle breeds, which confirm the importance of such loci for the trait variation. In addition, novel genomic loci were suggested. Future validation studies with larger sample size should be done to verify the results obtained from the current study in order to fine mapping the identified genomic loci, which may play a role to increase the rate of genetic improvement for milk production traits in Egyptian buffalo using genomic approaches.

DECLARATIONS

Authors' Contribution

MAAA collected data and samples, contributed to analyses, and wrote the manuscript. SA-B contributed to reagents and materials preparation as well as data collection. HE-R collected data and samples and contributed to analyses. SE-A contributed to reagents and materials preparation along with genotyping. HA conceived and designed the experiment, analyzed the data and contributed to the manuscript writing.

Competing interests

The authors have not declared any conflict of interests.

To cite this paper: Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal S.E-D and Abdel-Shafy H (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo. *World Vet. J.*, 10 (1): 35-42. DOI: https://dx.doi.org/10.36380/scil.2020.wvj5

Acknowledgements

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".

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. DOI: https://doi.org/10.1016/j.livsci.2020.103977.

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. DOI: https://doi.org/10.1186/1297-9686-46-35.

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. DOI: http://doi.org/10.1038/nprot.2010.116.

Banos G, Bramis G, Bush SJ, Clark EL, McCulloch MEB, Smith J, Schulze G, Arsenos G, Hume DA and Psifidi A (2017). The genomic architecture of mastitis resistance in dairy sheep. BMC Genomics, 18: 624. DOI: https://doi.org/10.1186/s12864-017-3982-1.

Benedet A, Ho PN, Xiang R, Bolormaa S, De Marchi M, Goddard ME and Pryce JE (2019). The use of mid-infrared spectra to map genes affecting milk composition. Journal of Dairy Science, 102(8):7189-7203. DOI: https://doi.org/10.3168/jds.2018-15890.

Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, Herd RM and Goddard ME (2011). Genome-wide association studies for feedlot and growth traits in cattle. Journal of Animal Science, 89:1684-1697. DOI: 10.2527/jas.2010-3079.

Bouwman AC, Bovenhuis H, Visker MH and van Arendonk JA (2011). Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genetics, 12:43. DOI: 10.1186/1471-2156-12-43.

Buckley F, O'Sullivan K, Mee JF, Evans RD and Dillon P (2003). Relationships among milk yield, body condition, cow weight, and reproduction in spring-calved Holstein-Friesians. Journal of Dairy Science, 86: 2308-2319. DOI: https://10.3168/jds.S0022-0302(03)73823-5.

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. DOI: https://10.1186/s13742-015-0047-8.

Chen X, Cheng Z, Zhang S, Werling D and Wathes DC (2015). Combining Genome Wide Association Studies and Differential Gene Expression Data Analyses Identifies Candidate Genes Affecting Mastitis Caused by Two Different Pathogens in the Dairy Cow. Open Journal of Animal Sciences, 5: 358-393. DOI: https://10.4236/ojas.2015.54040.

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. DOI: https://doi.org/10.1186/s12864-015-1986-2.

Doran AG, Berry DP and Creevey CJ (2014). Whole genome association study identifies regions of the bovine genome and biological pathways involved in carcass trait performance in Holstein-Friesian cattle. BMC Genomics, 15: 837. DOI: https://10.1186/1471-2164-15-837.

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. DOI: https://doi.org/10.1016/j.livsci.2017.01.019.

Fang L, Hou Y, An J, Li B, Song M, Wang X, Sørensen P, Dong Y, Liu C, Wang Y, Zhu H, Zhang S and Yu Y (2016). Genome-wide transcriptional and post-transcriptional regulation of innate immune and defense responses of bovine mammary gland to staphylococcus aureus. Frontiers in Cellular and Infection Microbiology, 6:193. DOI: https://doi.org/10.3389/fcimb.2016.00193.

Finlay EK, Berry DP, Wickham B, Gormley EP and Bradley DG (2012). A genome wide association scans of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. PLoS One, 7(2): e30545. DOI: https://doi.org/10.1371/journal.pone.0030545.

Gantner V, Mijic P, Baban M, Skrtic Z and Turalija A (2015). The overall and fat composition of milk of various species. Mljekarstvo, 65: 223-231. DOI: https://10.15567/mljekarstvo.2015.0401.

Gonzalez Guzman JL, Lázaro SF, do Nascimento AV, de Abreu Santos DJ, Cardoso DF, Becker Scalez DC, Galvão de Albuquerque L, Hurtado Lugo NA, and Tonhati H (2020). Genome-wide association study applied to type traits related to milk yield in water buffaloes (Bubalus bubalis). Journal of Dairy Science, 103(2): 1642-1650. DOI: https://10.3168/jds.2019-16499.

Guo J, Jorjani H and Carlborg Ö (2012). A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. BMC Genetics, 13:82. DOI: https://doi.org/10.1186/1471-2156-13-82.

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. Page 825 in Proc. of the 11th World Congress on Genetics Applied to Livestock Production, New Zealand. Available at: http://www.wcgalp.org/proceedings/2018.

Hill WG (2012). Quantitative genetics in the genomics era. Current Genomics, 13(3):196-206. DOI: https://10.2174/138920212800543110.

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. DOI: https://10.1093/nar/gky1084.

Huang W, Kirkpatrick BW, Rosa GJ and Khatib H (2010). A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in Holstein cattle. Journal of Animal Genetics, 41:570-578. DOI: https://10.1111/j.1365-2052.2010.02046.x.

Iamartino D, Williams JL, Sonstegard T, Reecy J, Van Tassell C, Nicolazzi EL, Biffani S, Biscarini F, Schroeder S, de Oliveira DAA et al. (2013). The buffalo genome and the application of genomics in animal management and improvement. Buffalo Bulletin, 32:151-158. Avilable at: http://ibic.lib.ku.ac.th/e-Bulletin/IBBUSI201301012.pdf.

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. DOI: https://10.1186/s12863-016-0363-8.

Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, Boehm F, Caporaso NE, Cornelis MC, Edenberg HJ et al. (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. Genetic Epidemiology,34: 591-602. DOI: https://10.1002/gepi.20516.

Lin Y, Sun X, Hou X, Qu B, Gao X and Li Q (2016). Effects of glucose on lactose synthesis in mammary epithelial cells from dairy cow. BMC Veterinary Research, 12:81. DOI: https://10.1186/s12917-016-0704-x.

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. DOI: https://10.3168/jds.2017-13246.

To cite this paper: Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal S.E-D and Abdel-Shafy H (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo. *World Vet. J.*, 10 (1): 35-42. DOI: https://dx.doi.org/10.36380/scil.2020.wvj5

Mao Y, Zhu X, Xing S, Zhang M, Zhang H, Wang X, Karrow N, Yang L and Yang Z (2015). Polymorphisms in the promoter region of the bovine lactoferrin gene influence milk somatic cell score and milk production traits in Chinese Holstein cows. Research in Veterinary Science, 103: 107-112. DOI: https:// 10.1016/j.rvsc.2015.09.021.

Ménard O, Ahmad S, Rousseau F, Briard-Bion V, Gaucheron F and Lopez C (2010). Buffalo vs. cow milk fat globules: Size distribution, zeta-potential, compositions in total fatty acids and in polar lipids from the milk fat globule membrane. Food Chemistry, 120: 544-551. DOI: https://doi.org/10.1016/j.foodchem.2009.10.053.

Miglior F, Sewalem A, Jamrozik J, Lefebvre DM and Moore RK (2006).Analysis of milk urea nitrogen and lactose and their effect on longevity in Canadian dairy cattle. Journal of Dairy Science, 89: 4886-4894. DOI: https://www.ncbi.nlm.nih.gov/pubmed/17106119.

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T and Lee, D (2002). BLUPF90 and related programs (BGF90). 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France. Available at: http://www.wcgalp.org/proceedings/2002.

Mokhber M, Shahrbabak MM, Sadeghi M, Shahrbabak HM, Stella A, Nicolzzi E and Williams JL (2019). Study of whole genome linkage disequilibrium patterns of Iranian water buffalo breeds using the Axiom Buffalo Genotyping 90K Array. PLoS One, 14(5): e0217687. DOI: https://10.1371/journal.pone.0217687.

Nayeri S, Sargolzaei M, Abo-Ismail MK, May N, Miller SP, Schenkel F, Moore SS and Stothard P (2016). Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genetics, 17:75. DOI: https://doi.org/10.1186/s12863-016-0386-1.

Power RA, Parkhill J and de Oliveira T (2016). Microbial genome-wide association studies: lessons from human GWAS. Nature Reviews, 1:10. DOI: 10.1038/nrg.2016.132.

Price AL, Zaitlen NA, Reich D and Patterson N (2010). New approaches to population stratification in genome-wide association studies. Nature Reviews, 11:459-463. DOI: https://10.1038/nrg2813.

Ptak EPB and Bieniek J (2012). Genetic parameters for lactose percentage in the milk of Polish Holstein-Friesians. Journal of Animal and Feed Sciences, 21: 251-262. DOI: https://doi.org/10.22358/jafs/66072/2012.

Reist M, Erdin D, von Euw D, Tschuemperlin K, Leuenberger H, Chilliard Y, Hammon H M, Morel C, Philipona C, Zbinden Y, Kuenzi N and Blum JW (2002). Estimation of energy balance at the individual and herd level using blood and milk traits in high-yielding dairy cows. Journal of Dairy Science, 85: 3314-3327. DOI: https://doi.org/10.3168/jds.S0022-0302(02)74420-2.

Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, Bhattarai EK, Molaei M, Freetly HC, Hansen SL et al. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. BMC Genomics, 18:386. DOI: https://doi.org/10.1186/s12864-017-3754-y.

Szyda J, Mielczarek M, Frąszczak M, Minozzi G, Williams JL and Wojdak-Maksymiec K (2019). The genetic background of clinical mastitis in Holstein Friesian cattle. Animal, 13(10): 2156 -2163. DOI: https://10.1017/S1751731119000338.

Thermo Fisher Scientific. Axiom™ Buffalo Genotyping Array. Accessed on June 2019. Available at: https://www.thermofisher.com/order/catalog/product/550431.

VanRaden PM and Wiggans GR (1991). Derivation, calculation, and use of national animal model information. Journal of Dairy Science, 74: 2737-2746. DOI: https://doi.org/10.3168/jds.S0022-0302(91)78453-1.

Venturini GC, Cardoso DF, Baldi F, Freitas AC, Aspilcueta-Borquis RR, Santos DJ, Camargo GM, Stafuzza NB, Albuquerque LG and Tonhati H (2014). Association between single-nucleotide polymorphisms and milk production traits in buffalo. Genetics and Molecular Research. 13: 10256-10268. DOI: https://10.4238/2014.

Wang S, Dvorkin D and Da Y (2012). SNPEVG: a graphical tool for GWAS graphing with mouse clicks. BMC Bioinformatics, 13:319. DOI: https://10.1186/1471-2105-13-319.

Welderufael BG, Løvendahl P, de Koning D-J, Janss LLG and Fikse WF (2018). Genome-Wide Association Study for Susceptibility to and Recoverability From Mastitis in Danish Holstein Cows. Frontiers in Genetics, 9:141. DOI: https://10.3389/fgene.2018.00141.

Wu X, Fang M, Liu L, Wang S, Liu J, Ding X, Zhang S, Zhang Q, Zhang Y, Qiao L, Lund MS, Su G and Sun D (2013). Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. BMC Genomics, 14:897. DOI: https://10.1186/1471-2164-14-897.

To cite this paper: Awad MAA, Abou-Bakr S, El-Regalaty H, El-Assal S.E-D and Abdel-Shafy H (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian Buffalo. *World Vet. J.*, 10 (1): 35-42. DOI: https://dx.doi.org/10.36380/scil.2020.wvj5