

Article



# **Genome-Wide Association Study of Lactation Traits in Chinese** Holstein Cows in Southern China

Minqiang Su<sup>1,2,3,†</sup>, Xiaojue Lin<sup>1,2,3,†</sup>, Zupeng Xiao<sup>1,2,3</sup>, Yuanhang She<sup>1,2,3</sup>, Ming Deng<sup>1,2,3</sup>, Guangbin Liu<sup>1,2,3</sup>, Baoli Sun<sup>1,2,3</sup>, Yongqing Guo<sup>1,2,3</sup>, Dewu Liu<sup>1,2,3,\*</sup> and Yaokun Li<sup>1,2,3,\*</sup>

- <sup>1</sup> College of Animal Science, South China Agricultural University, Guangzhou 510642, China; smq2054113415@163.com (M.S.); jasonlin1996@163.com (X.L.); xiaozupeng@stu.scau.edu.cn (Z.X.); syh15521170780@163.com (Y.S.); dengming@scau.edu.cn (M.D.); gbliu@scau.edu.cn (G.L.); baolisun@scau.edu.cn (B.S.); yongqing@scau.edu.cn (Y.G.)
- <sup>2</sup> National Local Joint Engineering Research Center of Livestock and Poultry, South China Agricultural University, Guangzhou 510642, China
- <sup>3</sup> Guangdong Key Laboratory of Agricultural Animal Genomics and Molecular Breeding, South China Agricultural University, Guangzhou 510640, China
- \* Correspondence: dwliu@scau.edu.cn (D.L.); ykli@scau.edu.cn (Y.L.)
- <sup>+</sup> These authors contributed equally to this work.

**Simple Summary:** Lactation traits are economically important traits in dairy production, with milk yield, milk fat percentage, and milk protein percentage being the main indicators of cow performance and milk quality. The interaction between environment and genetics has a great influence on the performance of dairy cows. Cows are susceptible to heat stress under conditions of high temperature and humidity, which in turn affects lactation traits. Therefore, the aim of this study was to identify single-nucleotide polymorphism loci and candidate genes related to lactation traits in Holstein dairy cows under conditions of high temperature and high humidity in southern China. Our study identified 23 single-nucleotide polymorphic sites significantly associated with lactation traits, and 10 key candidate genes associated with lactation traits. The results of this study may provide a theoretical basis for the molecular marker breeding of Chinese Holstein cows in high-temperature and high-humidity environments in southern China.

Abstract: Lactation traits are economically important for dairy cows. Southern China has a hightemperature and high-humidity climate, and environmental and genetic interactions greatly impact dairy cattle performance. The aim of this study was to identify novel single-nucleotide polymorphism sites and novel candidate genes associated with lactation traits in Chinese Holstein cows under hightemperature and humidity conditions in southern China. A genome-wide association study was performed for the lactation traits of 392 Chinese Holstein cows, using GGP Bovine 100 K SNP gene chips. Some 23 single nucleotide polymorphic loci significantly associated with lactation traits were screened. Among them, 16 were associated with milk fat rate, 7 with milk protein rate, and 3 with heat stress. A quantitative trait locus that significantly affects milk fat percentage in Chinese Holstein cows was identified within a window of approximately 0.5 Mb in the region of 0.4–0.9 Mb on Bos taurus autosome 14. According to Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses, ten genes (DGAT1, IDH2, CYP11B1, GFUS, CYC1, GPT, PYCR3, OPLAH, ALDH1A3, and NAPRT) associated with lactation fat percentage, milk yield, antioxidant activity, stress resistance, and inflammation and immune response were identified as key candidates for lactation traits. The results of this study will help in the development of an effective selection and breeding program for Chinese Holstein cows in high-temperature and humidity regions.

**Keywords:** genome-wide association study; Chinese Holstein cows; high temperature and humidity; lactation traits



Citation: Su, M.; Lin, X.; Xiao, Z.; She, Y.; Deng, M.; Liu, G.; Sun, B.; Guo, Y.; Liu, D.; Li, Y. Genome-Wide Association Study of Lactation Traits in Chinese Holstein Cows in Southern China. *Animals* **2023**, *13*, 2545. https://doi.org/10.3390/ ani13152545

Academic Editor: Andrea Criscione

Received: 25 June 2023 Revised: 4 August 2023 Accepted: 6 August 2023 Published: 7 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

# 1. Introduction

Consumer demand for high-quality dairy products is increasing with improvements in the living standards of the population. Lactation traits are very important economic traits in dairy production [1]. Lactation traits such as milk yield (MY), milk fat percentage (FP), and milk protein percentage (PP) are the main indicators of cow productivity and milk quality [2,3]. Changes in the environment, especially high temperature and humidity, have a significant impact on cow production performance [4]. Southern China has a subtropical climate with long periods of high temperatures and high humidity, and cows are subjected to heat stress in such environments. The effects of heat stress on dairy production can be categorized into two different causes: effects caused by reduced voluntary feed intake due to heat stress, and direct physiological and metabolic effects of heat stress [5]. Heat stress reduces MY and alters FP and PP through both thus [6,7]. Heat stress has been shown to affect the expression of genes related to milk fat metabolism [8]. Genomic regions associated with heat tolerance in dairy cows (Bos taurus autosome (BTA)5, BTA14 and BTA15) were identified by Sigdel et al. [9].

Genome-wide association study (GWAS) plays an important role in dairy breeding by helping to identify key genes, speeding up the breeding process and improving the genetic improvement of target traits, and many candidate genes and variant sites associated with lactation traits have been identified in Chinese Holstein cows by using GWAS. GWAS analysis of five milk yield traits in a Chinese Holstein cow population was performed by Jiang et al. [10]. A total of 105 significant single nucleotide polymorphism (SNP) loci were screened and localized to *DGAT1*, *GHR*, *CYP11B1*, *ABCG2*, and *LGB*. A GWAS was performed by Iung et al. [11] using 56,256 SNPs from 1067 cows to screen eight genes significantly associated with lactation traits. However, few GWAS have been reported for Chinese Holstein cows under high temperature and humidity conditions. Therefore, the aim of this study was to identify SNP sites and candidate genes related to lactation traits in Chinese Holstein dairy cows under high temperature and humidity conditions and provide a theoretical basis for molecular marker breeding of Chinese Holstein cows in high temperature and humidity environments.

#### 2. Materials and Methods

### 2.1. Animal and Phenotypic Data

The Chinese Holstein cows used in the study were obtained from a dairy farm in Hezhou, Guangxi, China. The local climate is subtropical monsoon, with average annual temperatures ranging from 19.2 to 19.9 °C and average annual precipitation of 1558.1 to 2012.1 mm. All animals were raised under the same conditions, and healthy cows without diseases are screened according to the records of cattle diseases of pasture veterinarians. The experimental population consisted of 392 Chinese Holstein cows, and the pedigree of the cows can be traced back at least three generations. MY data were farm milk hall statistics. (For cows in the current lactation period with a lactation duration exceeding 305 days, the 305-day milk yield was determined based on actual statistical data. Conversely, for cows with a lactation duration less than 305 days, the 305-day milk yield was estimated using lactation curve analysis.) FP and PP were obtained by averaging five measurements after milk samples were collected by the farm each month and sent to Dairy Herd Improvement Center (DHI, China) and analyzed in the laboratory.

## 2.2. Genotyping and Quality Control

Blood samples were collected from the experimental cows, and DNA was extracted from the blood samples by using a Magnetic Universal Genomic DNA Kit (TIANGEN, Beijing, China). Genotyping was performed using the GGP Bovine 100 K SNP GeneChip on the Illumina iScan SNP microarray typing platform. Quality control was performed using the PLINK software (v1.9) [12], and the quality control criteria were (1) exclusion of SNPs at unknown locations and on sex chromosomes, (2) exclusion of individuals with SNP deletions greater than 5%, (3) exclusion of SNPs with detection rates less than 95%,

(4) exclusion of SNPs with sub-allele frequencies less than 0.05, and (5) exclusion of SNPs that deviated from Hardy–Weinberg equilibrium ( $p < 1.0 \times 10^{-6}$ ).

## 2.3. Phenotypic Data Processing

Phenotypic data for lactation traits were collated. Abnormal values that deviated from the phenotypic mean  $\pm$  3 times the standard deviation were excluded. Fixed-effect levels were classified as follows: calving age was divided into six levels, i.e., <24, 25–36, 37–48, 49–60, 61–72, and >72 months; parity was divided into five levels, from the first to the fifth parity; calving season was divided into four levels, November–February, March–April, September–October, and May–August according to the average maximum temperature of January–December in Guangxi Hezhou City.

### 2.4. Fixed-Effects ANOVA

A general linear model with the ANOVA function of R 4.1 software was used to perform ANOVA of the fixed effects. Three effects of calving month, calving season, and litter size were considered for lactation traits, and the model was as follows:

$$Y_{ijkl} = \mu + Month_i + Season_j + Parity_k + e_{ijkl}$$

where  $Y_{ijkl}$  is the observed value of lactation traits;  $\mu$  is the population mean; *Month<sub>i</sub>* is the calving age effect; *Season<sub>j</sub>* is the calving season effect; *Parity<sub>k</sub>* is the parity effect; and  $e_{ijkl}$  is the residual.

## 2.5. Estimation of Individual Breeding Values

The single-trait animal model of the DMUAI module of the derivative-free approach to multivariate analysis (DMU, Version 6) software was used to estimate the variance components of lactation traits for all individuals in the test population, and the variancecovariances of the obtained fixed effects were used as a priori values to calculate the estimated breeding values of the individuals by using the DMU4 module. The models for estimating the variance components of lactation traits were as follows:

$$Y = Xz + Wu + e$$

where Y is the observed value, z is the fixed effect (calving age, calving season, and parity), X is the correlation matrix of fixed effects, u is the random effect (individual ID), W is the correlation matrix of random effects, and e is the residual.

### 2.6. Population Structure and Kinship Identification

Principal component analysis (PCA) was performed using Tassel 5.2.8 software to examine the existence of stratification in the test population. Based on the first three principal components of the PCA results, the population structure map was drawn using the ggplot package of R 4.1 software. On the basis of the population structure map, we observed whether the population was stratified and added the results of the first three principal components to the correlation analysis of the mixed linear model. The kinship matrix was constructed using Tassel 5.2.8 software with genotype data and visualized using R 4.1 software, and a kinship heat map was drawn to observe the existence of kinship between individuals. The kinship matrix results were added to the mixed linear model for association analysis.

## 2.7. Genome-Wide Association Study

In this study, individual estimated breeding values were used as phenotypes with Tassel 5.2.8 software and correlated with SNP gene chip typing data for analysis. The mixed linear model used in this experiment was as follows:

$$Y = S\alpha + Q\nu + Zu + e$$

where *Y* is the individual breeding value of lactation traits;  $\alpha$  is the fixed effect of SNP markers; *v* is the population structure effect; *u* is the residual polygenic effect,  $u \sim N (0.2K\sigma_a^2)$ , where *K* is the kinship matrix and  $\sigma_a^2$  is the genetic variance; *e* is the residual; and *S*, *Q*, and *Z* are the association matrices of  $\alpha$ , *v*, and *u*, respectively. The results obtained from the association analysis were corrected for multiple hypothesis testing using the Bonferroni method. Briefly, N (81,203) was the total number of filtered SNPs, and the genomic level significance (6.1574 × 10<sup>-7</sup>, 0.05/N) and chromosome level significance (1.2315 × 10<sup>-5</sup>, 1/N) were used as thresholds.

## 2.8. Identification of Candidate Genes and Functional Enrichment Analysis

In this study, Ensembl UMD3.1 (http://asia.ensembl.org/Bos\_taurus/Info/Index/, accessed on 2 October 2021) database was used to locate genes within 1 Mb upstream and downstream of significant SNP sites. To better understand the biological processes, gene ontology (GO) terms and encyclopedia of genes and genomes (KEGG) pathways were annotated, visualized, and integrated for discovery on the basis of relevant genes by using the DAVID 6.8 database (https://david.ncifcrf.gov/, accessed on 18 October 2021). A *p*-value < 0.05 was the threshold for significant enrichment of GO terms and KEGG pathways.

## 3. Results

#### 3.1. Descriptive Statistics and Estimated Breeding Values of Lactation Ttraits

Table 1 summarizes the descriptive statistics for the three lactation traits, MY, FP, and PP, of the test population with means of 9807.07, 3.54, and 3.19, respectively, and coefficients of variation (CV) of 15.64%, 13.56%, and 7.21%, respectively. Table 2 shows the results of the ANOVA for lactation trait fixed effects. From the results, it is clear that the calving age effect had a highly significant effect on MY and FP (p < 0.001) and no significant effect on PP. The parity effect had no significant effect on MY and PP and a significant effect on FP (p < 0.01). The calving season effect had a highly significant effect on both MY (p < 0.001) and FP (p < 0.01) and no significant effect on PP. Therefore, correction for calving age and calving season effects was required to estimate MY individual breeding values, and correction for calving age, parity, and calving season effects was performed for PP individual breeding values. The statistical results of the estimated breeding values for the lactation traits MY, FP, and PP are shown in Table 3, with means of 70.25, -0.03, and -0.01 and estimated heritability of 0.07, 0.12, and 0.20, respectively.

**Table 1.** Descriptive statistics of lactation traits <sup>1</sup>.

Traits	п	Mean	Min	Max	SD	CV (%)
MY	392	9807.07	4870.42	14205.58	1538.60	15.64
FP	392	3.54	1.97	5.07	0.48	13.56
PP	392	3.19	2.34	3.76	0.23	7.21

<sup>1</sup> MY, milk yield; FP, milk fat percentage; PP, milk protein percentage. CV, coefficient of variation.

**Table 2.** Analysis of variance components of lactation traits phenotypes <sup>1</sup>.

Item	МҮ	FP	РР
Calving age	<0.001 ***	<0.001 ***	0.797
Parity	0.101	0.002 **	0.296
Calving season	<0.001 ***	0.007 **	0.054

<sup>1</sup> MY, milk yield; FP, milk fat percentage; PP, milk protein percentage. \*\* indicates significant at the p < 0.01 level, and \*\*\* indicates significant at the p < 0.001 level.

Traits	п	Mean	Min	Max	SD	Heritability
MY	392	70.25	-1013.32	1118.44	408.94	0.07
FP	392	-0.03	-0.45	0.41	0.16	0.12
PP	392	-0.01	-0.34	0.26	0.10	0.20

**Table 3.** Statistics of the estimated breeding values for lactation traits <sup>1</sup>.

<sup>1</sup> MY, milk yield; FP, milk fat percentage; PP, milk protein percentage.

#### 3.2. Population Stratification

The PCA results (Figure 1A) showed that the test population was divided into several subpopulations of different sizes with significant stratification. Therefore, the results of PCA were included as covariates in the association analysis model and population stratification based on the results of PCA analysis was considered in the mixed linear model. The kinship mapping is shown in Figure 1B. Many individuals in the test population were related to each other, and there were eight larger families consisting of many related individuals. Therefore, the kinship of individuals within the test population needed to be considered in the association analysis, and the effect of kinship between individuals on the association analysis results was corrected by including the kinship matrix as a random effect in the mixed linear model of association analysis.



Figure 1. (A) The PCA plot of top two principal components. (B) The kinship matrix diagram.

#### 3.3. Genome-Wide Association Results

The Manhattan plot and quantile-quantile plot (Q-Q plot) of the association analysis are shown in Figure 2. According to the Manhattan plot, numerous SNP loci significantly associated with the FP trait were located on BTA14. These significant loci form peaks of SNP loci, as observed from the results. Combined with Table 4, it can be inferred that these significant loci were mainly situated within a 0.5 Mb window ranging from 0.4 to 0.9 Mb on BTA14, with a total of 15 significant SNP loci, which was the smallest range and the most concentrated distribution of SNP loci in this association analysis. Based on the principle of linkage disequilibrium, it is possible that some of these significant SNP loci were false positives, showing strong linkage disequilibrium with the true associated SNP loci, leading to significant results. Therefore, it is necessary to comprehensively discuss these loci considering their physical positions, associated genes, and previous studies. Regarding the SNP loci significantly associated with the PP trait, they primarily resided in a 2 Mb window on BTA11 (3 SNPs). On the other hand, the three significant SNP loci on BTA21 were relatively scattered. No significant SNP loci associated with the MY trait were identified in this study. Observing the Manhattan plot and Q-Q plot of the MY trait, it can be observed that the Q-Q plot of the MY trait exhibited a good fit. This may be attributed to the conservative correction imposed by the Bonferroni method in multiple hypothesis

testing, which could filter out some SNP loci that might have significant associations with the MY trait, particularly for large-scale chip data. Another possibility is that the lack of association between the MY trait and significant SNP loci could be due to the relatively small sample size in the association analysis.



**Figure 2.** Manhattan plots and Q–Q plots from GWAS for MY, FP, and PP traits in Chinese Holstein cows. (A) MY: milk yield; (B) FP: milk fat percentage; (C) PP: milk protein percentage. In the Manhattan plot, above the blue horizontal line indicates a significant SNP chromosome level and above the red horizontal line indicates a significant SNP genomic level.

Table 4 results demonstrate that a total of 23 SNP loci were found to be significantly associated with lactation traits in this study, with 13 SNP loci showing significance at the genome level ( $p < 6.1574 \times 10^{-7}$ ) and 10 SNP loci exhibiting significance at the chromosome level ( $p < 1.2315 \times 10^{-5}$ ). Among the SNP loci associated with FP trait, there were a total of 16 loci, with 15 of them located on BTA14. The most significantly associated SNP with the FP trait was identified as ARS-BFGL-NGS-4939, which is situated within the diacylglycerol O-acyltransferase 1 gene (*DGAT1*). Within the *PLEC* gene, two SNP loci were found to be significantly associated with the FP trait. For the PP trait, a total of 7 SNP loci were significantly associated, including 3 loci on BTA11, all located within the *SNRNP200* gene, and 3 loci on BTA21, respectively, within the *CERS3*, *AP3S2*, and *GABRA5* genes. The most significantly associated SNP with the MP trait was identified as Hapmap32084-BTA-147824, which is situated within the *CERS3* gene.

Traits	SNP Name	BTA	Position (bp)	<i>p</i> -Value	Nearest Gene Name	Distance (bp)
	ARS-BFGL-NGS-4939	14	609,870	$2.85 imes10^{-9}$	DGAT1	Within
	Chr14_1765835	14	580,019	$2.96 imes10^{-9}$	SLC52A2	Within
	BovineHD1400000216	14	550,784	$3.12  imes 10^{-9}$	CPSF1	Within
	Chr14_1757935	14	572,120	$2.74 imes10^{-8}$	ADCK5	1622
	Chr14_2022745	14	831,004	$6.44 imes10^{-8}$	GRINA	938
	Chr14_1699016	14	513,203	$8.36 imes10^{-8}$	VPS28	384
	BovineHD1400000282	14	859,251	$9.51 imes10^{-8}$	PLEC	Within
FP	BovineHD1400000287	14	883,732	$9.51 imes10^{-8}$	PLEC	Within
	BovineHD1400000275	14	2,019,390	$1.00 \times 10^{-7}$	TSNARE1	Within
	ARS-BFGL-NGS-57820	14	465,742	$1.46 imes10^{-7}$	FOXH1	3390
	BovineHD1400000206	14	494,621	$2.25 \times 10^{-7}$	TONSL	1650
	Chr14_1653693	14	468,124	$2.52 \times 10^{-7}$	FOXH1	1008
	BovineHD1400000204	14	487,527	$3.05 imes10^{-7}$	CYHR1	Within
	DB-813-seq-rs110906821	14	758,854	$8.04 imes10^{-7}$	EXOSC4	1844
	BovineHD1200022127	12	73,935,024	$2.02  imes 10^{-6}$	HS6ST3	Within
	BovineHD1400000271	14	810,116	$5.73 imes10^{-6}$	SPATC1	3725
	Hapmap32084-BTA-147824	21	6,095,419	$1.40 imes10^{-6}$	CERS3	Within
РР	Hapmap25132-BTA-96391	13	27,232,535	$2.37  imes 10^{-6}$	FZD8	445,190
	ARS-BFGL-NGS-4300	21	21,278,489	$3.94 imes10^{-6}$	AP3S2	Within
	ARS-BFGL-NGS-40264	11	2,382,261	$6.04 imes10^{-6}$	SNRNP200	Within
	BovineHD2100000735	21	4,285,575	$8.19 imes10^{-6}$	GABRA5	Within
	BovineHD1100000821	11	2,385,285	$9.53 imes10^{-6}$	SNRNP200	Within
	BovineHD1100000817	11	2,381,644	$1.10 imes10^{-5}$	SNRNP200	Within

Table 4. Genome-wide significant SNPs associated with FP and PP traits and nearest candidate genes.

## 3.4. GO and KEGG Analyses for Lactation Traits

We performed GO and KEGG pathway enrichment analysis on a total of 145 genes located within a 1 Mb range upstream and downstream of the significant SNP loci. GO analysis revealed a significant enrichment (p < 0.05) for a total of 45 GO terms, including 16 biological process terms, 4 cellular component terms, and 25 molecular function terms (Figure 3, Table S1 and Table 5). The KEGG pathways analysis revealed that three pathways were enriched. Among them, two pathways (bta01100: Metabolic pathways and bta00480: glutathione metabolism) were significantly enriched, whereas the third pathway (bta01230: biosynthesis of amino acids) did not have a significant *p*-value (p = 0.067; Table 4). *IDH2* was enriched in three pathways; *ANPEP*, *OPLAH*, *GPT*, and *PYCR3* were enriched in two pathways.



Figure 3. GO analysis of regional candidate genes for lactation traits.

Term	Description	Gene Count	%	<i>p</i> -Value	Gene
bta01100	Metabolic pathways	16	11.1	0.03333	CERS3, DGAT1, GPAA1, IDH2, GPT, PYCR3, OPLAH, HDDC3, ALDH1A3, MAN2A2, NAPRT, CYP11B1, ANPEP, CYC1, GPAT2, GFUS
bta00480	Glutathione metabolism	3	2.1	0.04989	ANPEP, IDH2, OPLAH
bta01230	Biosynthesis of amino acids	3	2.1	0.06663	IDH2, GPT, PYCR3

**Table 5.** KEGG analysis of regional candidate genes for lactation traits.

## 4. Discussion

High-temperature and -humidity environments can contribute to metabolic imbalances and cause heat stress in cows, who experience a range of physiological and behavioral responses under heat stress conditions [8]. Heat stress decreases dry matter intake, affects nutrient metabolic activity, and reduces MY, FP, and PP in cows [4,6]. Heat stress not only affects milk production and quality, it also affects the overall health of the cow, affecting normal physiology, metabolism, hormones, and the immune system [13]. In this study, GWAS analysis was performed for lactation traits in Holstein cows under high temperature and high humidity conditions in Southern China. The GWAS results obtained 23 SNP loci that were significantly associated with lactation traits. Of the significantly associated SNPs, 16 were associated with FP and 7 were associated with PP. No SNPs significantly associated with MY traits were obtained, which indicates the limitations of this study, such as the sample size of the test population and the restricted genotyping platform. Among the significantly associated SNPs, 11 loci were previously reported to be significantly associated with lactation traits. For example, a study found that Chr14\_1765835, Chr14\_1757935, Chr14\_2022745, and Chr14\_1653693 were significantly associated with MY and PP [14]. Other studies found that BovineHD1400000216, BovineHD1400000275, Bo-vineHD1400000206, BovineHD1400000204, and BovineHD1400000271 were significantly associated with MY, FP, and PP in dairy cows [15,16]. Among the 16 SNP sites significantly associated with FP, 14 SNP sites were located within a window of approximately 0.5 Mb between 0.4 and 0.9 Mb on BTA14, and 68 genes were associated within 1 Mb upstream and downstream of this window. Numerous studies have shown the presence of a large number of quantitative trait loci (QTLs) and variant sites that affect lactation traits in dairy cows in the range of 0.4 to 0.9 Mb on BTA14. A SNP locus (rs137205809) which significantly affects FP was found at 0.7 Mb of BTA14 [17]. A large number of SNP sites significantly related to FP were found at 0.7 Mb of BTA14 [18]. Pausch et al. [19] performed targeted sequence padding of four candidate causal variant locus regions and association analysis of milk components and successfully identified SNP sites at 0.6 Mb on BTA14 that were significantly associated with FP. These findings are very close to the results of the present study, so QTLs that significantly affect FP in Chinese Holstein cows can be hypothesized to exist within a window of approximately 0.5 Mb between 0.4 and 0.9 Mb on BTA14.

In this study, novel SNP loci associated with lactation traits were discovered and have not been previously reported. Among these loci, the most significant one was Chr14\_1699016 ( $p = 8.36 \times 10^{-8}$ ), which is located 384 bp away from *VPS28* on BTA14 and approximately 37 kb away from the previously reported BovineHD1400000216 locus. Some studies have reported that *VPS28* gene was highly expressed in breast tissue and significantly associated with FP [10,20]. The SNP loci BovineHD1400000282 and BovineHD1400000287 are located within the PLEC gene on chromosome 14, which has been reported to play a crucial role in lactation processes [21]. Furthermore, studies have suggested that the PLEC gene may be associated with disease resistance or alleviation of heat stress [22]. DB-813-seq-rs110906821 is located on BTA14 at 1844 bp away from the *EXOSC4* gene, which has been reported to be significantly associated with colostrum

and serum protein concentrations in Chinese Holstein cows [23]. BovineHD1200022127 is located within the *HS6ST3* gene on BTA12, a member of the gene family involved in heparin metabolism. Heparin decreases the rate of lipoprotein lipase degradation in adipocytes and promotes adipocyte differentiation [24]. ARS-BFGL-NGS-4300 is located within the *AP3S2* gene on BTA21. A previous study found that *AP3S2* expression was significantly upregulated in Sahiwal cattle at 4, 24, and 48 h after heat stress, suggesting that the *AP3S2* gene may be associated with heat stress [25]. BovineHD2100000735 is located within the *GABRA5* gene on BTA21. GABRA5 is a subunit of gamma-aminobutyric acid A receptor, which plays a crucial role in maintaining ion balance. Ion balance is vital for lactation in cows [26]. In conclusion, the three newly discovered heat stress-related SNP loci (BovineHD140000282, BovineHD140000287, and ARS-BFGL-NGS-4300) identified in this study are potentially associated with lactation traits in cows under heat stress conditions.

In this study, the biological processes, cellular components and molecular functions obtained from GO analysis did not directly affect lactation traits. KEGG analysis identified 16 genes (DGAT1, CYC1, CERS3, GPAA1, IDH2, GPT, PYCR3, OPLAH, HDDC3, ALDH1A3, MAN2A2, NAPRT, CYP11B1, ANPEP, GPAT2, and GFUS) that were enriched in three pathways, namely metabolic pathways, glutathione metabolism, and biosynthesis of amino acids. Metabolic pathways are crucial for lactation in cows [15], while glutathione plays a significant role in antioxidant stress response and nutrient metabolism [27]. Moreover, amino acid biosynthesis serves as an important source of protein in milk [28]. Numerous studies have reported the association between the *DGAT1* gene and milk fat synthesis [29,30]. Furthermore, the DGAT1 gene is a major gene in inflammatory response and lipid metabolism, encoding an enzyme that catalyzes the last step of triacylglycerol synthesis [31], and significantly affecting energy metabolism during lactation in dairy cows [32]. The CYC1 gene encodes a subunit of the cytochrome bc1 complex, which is associated with redox processes, and it has been reported that the CYC1 gene is an important candidate for albumin synthesis [23]. Albumin is known to have anti-inflammatory and antioxidant functions [33], so the CYC1 gene may be associated with anti-inflammatory responses. The *IDH2* gene encodes isocitrate dehydrogenase. In the absence of ATP-citrate lyase, ruminants can use isocitrate dehydrogenase to form NADPH, which raises energy for cells, and the excess energy is converted to fatty acids, which are then stored as fat [34]. GWAS analysis and pathway analysis of fat deposition traits in Nello cattle revealed the relationship between *IDH2* gene and fat deposition [35]. The *GPT* gene encodes glutamicpyruvate transaminase. Habeeb et al. [36] conducted a study comparing the physiological and biochemical changes in Egyptian native cows and introduced cows under hot summer conditions, finding that serum levels of glutamate aminotransferase were significantly higher in native cows than in introduced cows under hot conditions, suggesting that the GPT gene may be associated with heat stress in dairy cows. The PYCR3 gene encodes a protein belonging to the pyrroline-5-carboxylic acid reductase family, which has been shown to be involved in genotoxic, inflammatory, and oxidative stress responses [37,38]. The *OPLAH* gene encodes an enzyme involved in the  $\gamma$ -glutamyl cycle and is responsible for converting the degradation product of glutathione, 5-oxoproline, back to glutamate [39]. Since 5-oxoproline is an oxidative stress inducer [40], OPLAH functions as an anti-oxidative stress agent by removing this metabolite. Many studies have identified the OPLAH gene as a candidate gene for milk yield traits in Holstein cows [14,41]. The ALDH1A3 gene encodes the third enzyme of the aldehyde dehydrogenase family 1, which is involved in the conversion of retinaldehyde to retinoic acid and has detoxifying and antioxidant functions [42]. The NAPRT gene encodes nicotinic acid phosphate ribosyltransferase, which has a key role in the acute inflammatory response [43]. The CYP11B1 gene encodes steroid  $11-\beta$ -hydroxylase, which is involved in steroid hormone biosynthesis. Many studies have shown that the CYP11B1 gene is associated with milk yield and milk fat percentage in dairy cows [44]. The GFUS gene encodes GDP-L-amylose synthase, which is involved in inflammatory and immune responses by activating the immune network [45], and it has also been shown that GFUS is associated with milk protein percentage and milk albumin

content in dairy cows [23,46]. Heat stress affects the anti-inflammatory response and immune function of cows by reducing the production of antibodies, cytokines, chemokines, and heat shock proteins [47]. In addition, cows exposed to heat stress have shown reduced MY and FP as well as PP [6]. In this study, GWAS analysis of lactation traits in Chinese Holstein cows under high temperature and humidity in Southern China was performed to screen for genes reported to be directly related to lactation traits (*DGAT1*, *IDH2*, *CYP11B1*, and *GFUS*) and candidate genes not reported to be directly related to lactation traits but associated with antioxidant, anti-stress, inflammatory, immune response, and detoxification processes (*CYC1*, *GPT*, *PYCR3*, *OPLAH*, *ALDH1A3*, *NAPRT*, and *GFUS*).

# 5. Conclusions

In summary, this study screened 23 SNPs that were significantly associated with lactation traits in Chinese Holstein cows in the high temperature and humidity region of Southern China. Among them, three SNPs were associated with heat stress. In addition, this study identified QTLs that significantly affected milk fat percentage in Chinese Holstein cows within a window of approximately 0.5 Mb between 0.4 and 0.9 Mb on BTA14. Ten most promising candidate genes (*DGAT1*, *IDH2*, *CYP11B1*, *GFUS*, *CYC1*, *GPT*, *PYCR3*, *OPLAH*, *ALDH1A3*, and *NAPRT*) were revealed based on KEGG analysis. The results of this study are useful for molecular marker-assisted breeding of lactation traits in Chinese Holstein cows under high temperature and humidity regions to accelerate genetic improvement.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ani13152545/s1, Table S1: GO analysis of regional candidate genes for lactation traits.

Author Contributions: Conceptualization, M.S., X.L., D.L. and Y.L.; methodology, M.S. and X.L.; software, M.S.; validation, M.S., X.L. and Z.X.; formal analysis, M.S. and X.L.; investigation, M.S. and X.L.; resources, Z.X., Y.S., M.D., G.L., B.S. and Y.G.; data curation, M.S. and X.L.; writing—original draft preparation, M.S.; writing—review and editing, Z.X., Y.S., M.D., G.L., B.S. and Y.G.; visualization, M.S. and X.L.; supervision, D.L. and Y.L.; project administration, D.L. and Y.L.; funding acquisition, D.L. and Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Guangdong Provincial Key Research and Development Programme (2022B0202090003) and GuangDong Basic and Applied Basic Research Foundation-Enterprise (Wens) Joint Fund (2019B1515210017).

**Institutional Review Board Statement:** This study was approved by the Ethics Committee of South China Agricultural University (SCAU, Guangzhou, China; Approval number: 2021g018). All animals used in this study met the guidelines for the care and use of experimental animals established by the Ministry of Agriculture of China.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The variation data reported in this paper have been deposited in the genome variation map (GVM) in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, under accession number GVM000516 (https://bigd.big.ac.cn/gvm/getProjectDetail?Project=GVM000516, accessed on 8 April 2023).

Acknowledgments: We thank Wen's Hezhou Dairy Farm (Hezhou, China) for their assistance with our studies.

Conflicts of Interest: The authors declare no conflict of interest.

## References

- 1. Kahi, A.K.; Nitter, G. Developing breeding schemes for pasture based dairy production systems in Kenya: I. Derivation of economic values using profit functions. *Livest. Prod. Sci.* 2004, *88*, 161–177. [CrossRef]
- Qian, L.; Zhao, A.; Zhang, Y.; Chen, T.; Zeisel, S.H.; Jia, W.; Cai, W. Metabolomic Approaches to Explore Chemical Diversity of Human Breast-Milk, Formula Milk and Bovine Milk. *Int. J. Mol. Sci.* 2016, 17, 2128. [CrossRef]
- Chen, S.; Chai, M.; Tian, C.; Li, Y.; Deng, T.; Wu, H.; Liu, X. Genetic variants of fatty acid elongase 6 in Chinese Holstein cow. *Gene* 2018, 670, 123–129. [CrossRef]

- 4. West, J.W. Effects of Heat-Stress on Production in Dairy Cattle. J. Dairy. Sci. 2003, 86, 2131–2144. [CrossRef]
- Cowley, F.C.; Barber, D.G.; Houlihan, A.V.; Poppi, D.P. Immediate and residual effects of heat stress and restricted intake on milk protein and casein composition and energy metabolism. J. Dairy. Sci. 2015, 98, 2356–2368. [CrossRef]
- Gorniak, T.; Meyer, U.; Südekum, K.H.; Dänicke, S. Impact of mild heat stress on dry matter intake, milk yield and milk composition in mid-lactation Holstein dairy cows in a temperate climate. *Arch. Anim. Nutr.* 2014, 68, 358–369. [CrossRef]
- 7. Yang, L.; Yang, Q.; Yi, M.; Pang, Z.H.; Xiong, B.H. Effects of seasonal change and parity on raw milk composition and related indices in Chinese Holstein cows in northern China. *J. Dairy. Sci.* 2013, *96*, 6863–6869. [CrossRef]
- 8. Wang, D.; Chen, Z.; Zhuang, X.; Luo, J.; Chen, T.; Xi, Q.; Zhang, Y.; Sun, J. Identification of circRNA-Associated-ceRNA Networks Involved in Milk Fat Metabolism under Heat Stress. *Int. J. Mol. Sci.* **2020**, *21*, 4162. [CrossRef]
- 9. Sigdel, A.; Abdollahi-Arpanahi, R.; Aguilar, I.; Peñagaricano, F. Whole Genome Mapping Reveals Novel Genes and Pathways Involved in Milk Production Under Heat Stress in US Holstein Cows. *Front. Genet.* **2019**, *10*, 928. [CrossRef]
- Jiang, L.; Liu, J.; Sun, D.; Ma, P.; Ding, X.; Yu, Y.; Zhang, Q. Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS ONE* 2010, *5*, e13661. [CrossRef]
- Iung, L.H.S.; Petrini, J.; Ramírez-Díaz, J.; Salvian, M.; Rovadoscki, G.A.; Pilonetto, F.; Dauria, B.D.; Machado, P.F.; Coutinho, L.L.; Wiggans, G.R.; et al. Genome-wide association study for milk production traits in a Brazilian Holstein population. *J. Dairy. Sci.* 2019, 102, 5305–5314. [CrossRef]
- Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007, *81*, 559–575. [CrossRef]
- 13. Ranjitkar, S.; Bu, D.; Van Wijk, M.; Ma, Y.; Ma, L.; Zhao, L.; Shi, J.; Liu, C.; Xu, J. Will heat stress take its toll on milk production in China? *Clim. Chang.* 2020, *161*, 637–652. [CrossRef]
- 14. Wang, T.; Li, J.; Gao, X.; Song, W.; Chen, C.; Yao, D.; Ma, J.; Xu, L.; Ma, Y. Genome-wide association study of milk components in Chinese Holstein cows using single nucleotide polymorphism. *Livest. Sci.* **2020**, *233*, 103951. [CrossRef]
- 15. Nayeri, S.; Sargolzaei, M.; Abo-Ismail, M.K.; May, N.; Miller, S.P.; Schenkel, F.; Moore, S.S.; Stothard, P. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* **2016**, *17*, 75. [CrossRef]
- Nayeri, S.; Sargolzaei, M.; Abo-Ismail, M.K.; Miller, S.; Schenkel, F.; Moore, S.S.; Stothard, P. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. J. Dairy. Sci. 2017, 100, 1246–1258. [CrossRef]
- Wang, D.; Ning, C.; Liu, J.F.; Zhang, Q.; Jiang, L. Short communication: Replication of genome-wide association studies for milk production traits in Chinese Holstein by an efficient rotated linear mixed model. J. Dairy. Sci. 2019, 102, 2378–2383. [CrossRef]
- 18. Prakapenka, D.; Liang, Z.; Jiang, J.; Ma, L.; Da, Y. A Large-Scale Genome-Wide Association Study of Epistasis Effects of Production Traits and Daughter Pregnancy Rate in U.S. Holstein Cattle. *Genes* **2021**, *12*, 1089. [CrossRef]
- 19. Pausch, H.; Wurmser, C.; Reinhardt, F.; Emmerling, R.; Fries, R. Short communication: Validation of 4 candidate causative trait variants in 2 cattle breeds using targeted sequence imputation. *J. Dairy. Sci.* **2015**, *98*, 4162–4167. [CrossRef]
- Jiang, L.; Liu, X.; Yang, J.; Wang, H.; Jiang, J.; Liu, L.; He, S.; Ding, X.; Liu, J.; Zhang, Q. Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC Genom.* 2014, 15, 1105. [CrossRef]
- Jiang, J.; Ma, L.; Prakapenka, D.; VanRaden, P.M.; Cole, J.B.; Da, Y. A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. Front. Genet. 2019, 10, 412. [CrossRef] [PubMed]
- Yang, Z.; Lian, Z.; Liu, G.; Deng, M.; Sun, B.; Guo, Y.; Liu, D.; Li, Y. Identification of genetic markers associated with milk production traits in Chinese Holstein cattle based on post genome-wide association studies. *Anim. Biotechnol.* 2021, 32, 67–76. [CrossRef] [PubMed]
- 23. Lin, S.; Wan, Z.; Zhang, J.; Xu, L.; Han, B.; Sun, D. Genome-Wide Association Studies for the Concentration of Albumin in Colostrum and Serum in Chinese Holstein. *Animals* **2020**, *10*, 2211. [CrossRef] [PubMed]
- Jiang, Z.; Michal, J.J.; Wu, X.L.; Pan, Z.; MacNeil, M.D. The heparan and heparin metabolism pathway is involved in regulation of fatty acid composition. *Int. J. Biol. Sci.* 2011, 7, 659–663. [CrossRef]
- Mehla, K.; Magotra, A.; Choudhary, J.; Singh, A.K.; Mohanty, A.K.; Upadhyay, R.C.; Srinivasan, S.; Gupta, P.; Choudhary, N.; Antony, B.; et al. Genome-wide analysis of the heat stress response in Zebu (Sahiwal) cattle. *Gene* 2014, 533, 500–507. [CrossRef]
- 26. Silanikove, N.; Maltz, E.; Halevi, A.; Shinder, D. Metabolism of Water, Sodium, Potassium, and Chlorine by High Yielding Dairy Cows at the Onset of Lactation. J. Dairy. Sci. **1997**, 80, 949–956. [CrossRef]
- 27. Wu, G.; Fang, Y.Z.; Yang, S.; Lupton, J.R.; Turner, N.D. Glutathione metabolism and its implications for health. *J. Nutr.* 2004, 134, 489–492. [CrossRef]
- 28. Osorio, J.S.; Lohakare, J.; Bionaz, M. Biosynthesis of milk fat, protein, and lactose: Roles of transcriptional and posttranscriptional regulation. *Physiol. Genom.* **2016**, *48*, 231–256. [CrossRef]
- 29. Thaller, G.; Krämer, W.; Winter, A.; Kaupe, B.; Erhardt, G.; Fries, R. Effects of DGAT1 variants on milk production traits in German cattle breeds. *J. Anim. Sci.* 2003, *81*, 1911–1918. [CrossRef]
- 30. Schennink, A.; Stoop, W.M.; Visker, M.H.; Heck, J.M.; Bovenhuis, H.; van der Poel, J.J.; van Valenberg, H.J.; van Arendonk, J.A. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. *Anim. Genet.* **2007**, *38*, 467–473. [CrossRef]

- Cases, S.; Smith, S.J.; Zheng, Y.W.; Myers, H.M.; Lear, S.R.; Sande, E.; Novak, S.; Collins, C.; Welch, C.B.; Lusis, A.J.; et al. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc. Natl. Acad. Sci. USA* 1998, 95, 13018–13023. [CrossRef] [PubMed]
- Bhattarai, D.; Dad, R.; Worku, T.; Xu, S.; Ullah, F.; Zhang, M.; Liang, X.; Den, T.; Fan, M.; Zhang, S. The functions and mechanisms of sequence differences of DGAT1 gene on milk fat synthesis between dairy cow and buffalo. *J. Dairy. Res.* 2020, *87*, 170–174. [CrossRef] [PubMed]
- Aragão, D.P.; da Silva Souza, B.; de Brito, T.V.; de Araújo Bastos Santana, L.; de Paiva Silva, R.M.; de Oliveira, A.P.; da Cunha Pereira, A.C.T.; Ferreira, G.P.; Dos Reis Barbosa, A.L.; de Oliveira, J.S. The anti-inflammatory and antinociceptive activity of albumins from Crotalaria retusa seeds. *Biomed. Pharmacother.* 2017, 93, 536–542. [CrossRef] [PubMed]
- 34. Metallo, C.M.; Gameiro, P.A.; Bell, E.L.; Mattaini, K.R.; Yang, J.; Hiller, K.; Jewell, C.M.; Johnson, Z.R.; Irvine, D.J.; Guarente, L.; et al. Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **2011**, *481*, 380–384. [CrossRef]
- Martins, R.; Machado, P.C.; Pinto, L.F.B.; Silva, M.R.; Schenkel, F.S.; Brito, L.F.; Pedrosa, V.B. Genome-wide association study and pathway analysis for fat deposition traits in nellore cattle raised in pasture-based systems. *J. Anim. Breed. Genet.* 2021, 138, 360–378. [CrossRef] [PubMed]
- 36. Marai, I.F.; Habeeb, A.A.; Farghaly, H.M. Productive, physiological and biochemical changes in imported and locally born Friesian and Holstein lactating cows under hot summer conditions of Egypt. *Trop. Anim. Health Prod.* **1999**, *31*, 233–243. [CrossRef]
- Kuo, M.L.; Lee, M.B.; Tang, M.; den Besten, W.; Hu, S.; Sweredoski, M.J.; Hess, S.; Chou, C.M.; Changou, C.A.; Su, M.; et al. PYCR1 and PYCR2 Interact and Collaborate with RRM2B to Protect Cells from Overt Oxidative Stress. *Sci. Rep.* 2016, *6*, 18846. [CrossRef]
- Phang, J.M.; Pandhare, J.; Liu, Y. The metabolism of proline as microenvironmental stress substrate. J. Nutr. 2008, 138, 2008s–2015s. [CrossRef]
- 39. Liu, Y.; Hyde, A.S.; Simpson, M.A.; Barycki, J.J. Emerging regulatory paradigms in glutathione metabolism. *Adv. Cancer Res.* 2014, 122, 69–101. [CrossRef]
- Pederzolli, C.D.; Sgaravatti, A.M.; Braum, C.A.; Prestes, C.C.; Zorzi, G.K.; Sgarbi, M.B.; Wyse, A.T.; Wannmacher, C.M.; Wajner, M.; Dutra-Filho, C.S. 5-Oxoproline reduces non-enzymatic antioxidant defenses in vitro in rat brain. *Metab. Brain Dis.* 2007, 22, 51–65. [CrossRef]
- Lu, X.; Arbab, A.A.I.; Abdalla, I.M.; Liu, D.; Zhang, Z.; Xu, T.; Su, G.; Yang, Z. Genetic Parameter Estimation and Genome-Wide Association Study-Based Loci Identification of Milk-Related Traits in Chinese Holstein. *Front. Genet.* 2021, *12*, 799664. [CrossRef] [PubMed]
- 42. Hsu, L.C.; Chang, W.C.; Shibuya, A.; Yoshida, A. Human stomach aldehyde dehydrogenase cDNA and genomic cloning, primary structure, and expression in Escherichia coli. *J. Biol. Chem.* **1992**, *267*, 3030–3037. [CrossRef] [PubMed]
- Audrito, V.; Messana, V.G.; Deaglio, S. NAMPT and NAPRT: Two Metabolic Enzymes With Key Roles in Inflammation. *Front.* Oncol. 2020, 10, 358. [CrossRef] [PubMed]
- 44. Wibowo, T.A.; Gaskins, C.T.; Newberry, R.C.; Thorgaard, G.H.; Michal, J.J.; Jiang, Z. Genome assembly anchored QTL map of bovine chromosome 14. *Int. J. Biol. Sci.* 2008, *4*, 406–414. [CrossRef] [PubMed]
- 45. Wang, L.; Huang, J.; Jiang, M.; Lin, H. Tissue-specific transplantation antigen P35B (TSTA3) immune response-mediated metabolism coupling cell cycle to postreplication repair network in no-tumor hepatitis/cirrhotic tissues (HBV or HCV infection) by biocomputation. *Immunol. Res.* **2012**, *52*, 258–268. [CrossRef] [PubMed]
- Buaban, S.; Lengnudum, K.; Boonkum, W.; Phakdeedindan, P. Genome-wide association study on milk production and somatic cell score for Thai dairy cattle using weighted single-step approach with random regression test-day model. *J. Dairy. Sci.* 2022, 105, 468–494. [CrossRef]
- 47. Bagath, M.; Krishnan, G.; Devaraj, C.; Rashamol, V.P.; Pragna, P.; Lees, A.M.; Sejian, V. The impact of heat stress on the immune system in dairy cattle: A review. *Res. Vet. Sci.* 2019, *126*, 94–102. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.